

Finance and Economics Discussion Series

Federal Reserve Board, Washington, D.C.

ISSN 1936-2854 (Print)

ISSN 2767-3898 (Online)

The Swaps Strike Back: Evaluating Expectations of One-Year Inflation

Anthony M. Diercks, Colin Campbell, Steve Sharpe, and Daniel Soques

2023-061

Please cite this paper as:

Diercks, Anthony M., Colin Campbell, Steve Sharpe, and Daniel Soques (2023). “The Swaps Strike Back: Evaluating Expectations of One-Year Inflation,” Finance and Economics Discussion Series 2023-061. Washington: Board of Governors of the Federal Reserve System, <https://doi.org/10.17016/FEDS.2023.061>.

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

The Swaps Strike Back: Evaluating Expectations of One-Year Inflation*

Anthony M. Diercks[†]
Federal Reserve Board

Colin Campbell[‡]
Federal Reserve Board

Steven A. Sharpe[§]
Federal Reserve Board

Daniel Soques[¶]
UNCW

September 2023

Abstract

This study examines the forecasting performance of inflation swaps and survey-based expectations for one-year inflation. Conducting this exercise helps determine if one set of expectations can provide a cleaner signal about future inflation. The study finds that, overall, inflation swaps more frequently provide better forecasts of future inflation. Previous studies that found poor performance of swaps were strongly influenced by liquidity issues during the financial crisis and the pandemic. When these periods are excluded, swaps have superior predictive ability. Our analysis suggests that combining the two expectations can lead to even better forecasts. The optimal static combination is roughly an equal weighting of swaps and surveys. Alternatively, a dynamic smooth-transition regime switching model can also lead to superior performance and provide a clearer signal on expectations of future inflation. Recently, this measure has implied the Federal Reserve is expected to be closer to its inflation target over the next year than the surveys would suggest.

JEL: E31, E37

Keywords: Inflation Expectations, Inflation Swaps, Surveys, Forecasting

*All errors remain our sole responsibility. Special thanks to Travis Berge, Andrew Meldrum, Barbara Rossi, Sarah Zubairy, members of the Monetary and Financial Market Analysis section, and participants at the Monetary Affairs Lunch Workshop at the Federal Reserve Board for helpful comments and suggestions. The views expressed herein are those of the authors and not necessarily those of the Board of Governors of the Federal Reserve System.

[†]Division of Monetary Affairs, Federal Reserve Board; <https://www.anthonydiercks.com>; Email: anthony.m.diercks@frb.gov

[‡]Division of Monetary Affairs, Federal Reserve Board

[§]Division of Research and Statistics, Federal Reserve Board

[¶]Department of Economics and Finance, University of North Carolina Wilmington

1 Introduction

Proper measurement of inflation expectations is paramount to policymakers and market participants. In the absence of a perfect measure, one can try to distinguish between different proxies for expectations based on their relative forecasting ability. For instance, if a market-based expectation is consistently beset by sizable risk premiums or liquidity issues, or if a survey-based measure is affected by its small sample size or respondents' behavioral biases, one would expect this to reduce forecast performance. On that note, we reassess and compare the forecasting efficacy (or perhaps, the lack thereof) for two prominent but very different measures of one-year inflation expectations, the consensus forecast from the Blue Chip Economic Indicators and the expectations gauged from the inflation swaps market.

To our knowledge, a direct comparison of these two important measures of one-year expectations has not been conducted for the most recent time period. Determining if one of these measures consistently outperforms the other would be of considerable interest, especially in light of persistent divergences that have been observed. For instance, as shown in Figure 1, the expectation implied by inflation swaps have been persistently below survey-based measures over the past year, a marked reversal from a comparison to early 2022: for instance, on November 8, 2022 the one-year inflation swap rate was 2.9% while the corresponding Blue Chip survey was 3.7%. In terms of the big picture, the inflation swaps seem to imply the Federal Reserve is expected to be closer to its inflation target over the next year than the surveys would suggest.

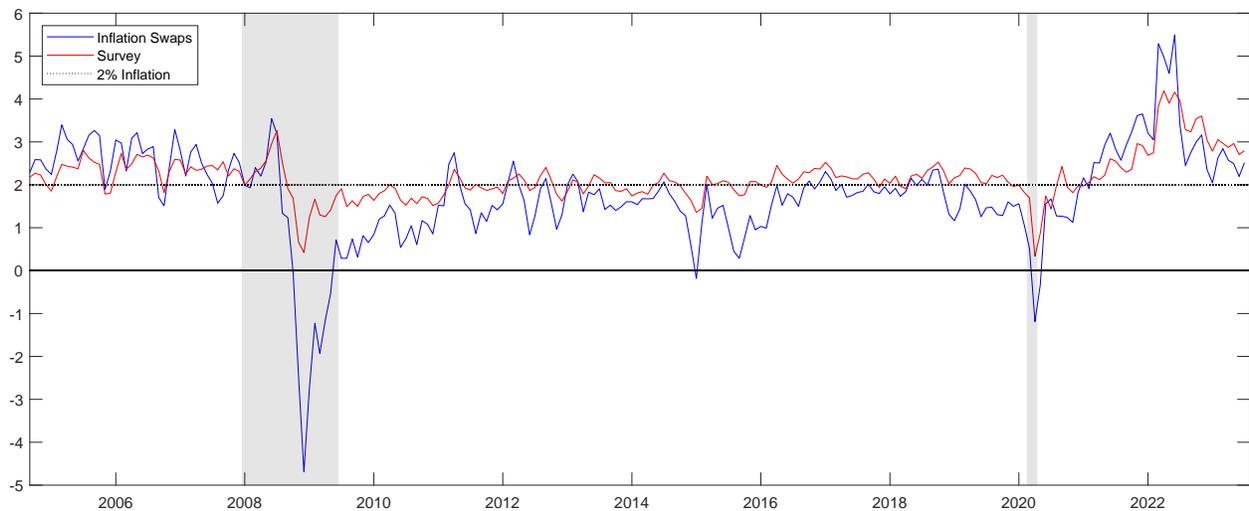


Figure 1: **One-Year Inflation Expectations of Inflation Swaps and Surveys**

Note: Gray bars denote NBER recessions. Source: Blue Chip Economic Indicators; Barclays, Bloomberg.

The gap between the surveys and market-based measures such as inflation swaps (see Figure 2) is typically believed to be due, in large part, to the presence of liquidity or risk premiums. These premiums represent the additional compensation or insurance that investors require for engaging in these trades. Thus, a widely held view is that surveys provide a cleaner measure of expectations, a factor that should ultimately translate into superior forecasting performance by survey forecasts. Consistent with this intuition, [Ang, Bekaert and Wei \(2007\)](#), [Gil-Alana, Moreno and Pérez de Gracia \(2012\)](#), [Faust and Wright \(2013\)](#), and [Bauer and McCarthy \(2015\)](#) all conclude that surveys generally outperform other measures in terms of forecasting.

That said, while survey-based expectations are not contaminated by risk premiums, they suffer from their own set of issues. For instance, surveys have often been found to be informationally inefficient ([Berge, 2018](#)), hampered by rigidities such as anchoring bias, forecast smoothing, as well as reputational concerns that induce forecasters to trade-off expected accuracy with stability and credibility.¹ Another characteristic that differentiates survey-based point forecasts from market-based forecasts is that the former tend to reflect the forecaster’s anticipated mode (i.e the most likely outcome) of the distribution, while market-based forecasts are more apt to reflect a probability-weighted mean.² In other words, the survey forecasts are likely to be influenced much less by changes in the likelihood of upper or lower tail outcomes because the mode excludes information on the shape of the rest of the distribution ([Diercks, Tanaka and Cordova, 2021](#)). Which measure one deems as superior depends on how one uses these forecasts. For instance, a good modal forecast should theoretically be closer to the realized value of inflation more frequently, but would tend to result in a larger mean squared error.

On that note, there have been several studies looking at somewhat more recent samples that find the surveys are no longer necessarily superior to other methods (see for example, [Trehan \(2015\)](#), [Kliesen \(2015\)](#), and [Berge \(2018\)](#)). Relatedly, [Diercks and Munir \(2020\)](#) show that at shorter horizons for interest rate expectations, surveys tend to underperform relative to expectations based on overnight index swap rates. These findings, along with the findings in [Diercks and Carl \(2019\)](#) and [Schmeling, Schrimpf and Steffensen \(2020\)](#) indicate that risk premiums at shorter horizons could generally be smaller and thus not as consequential as conventional wisdom would suggest.³ In other words, the signal from inflation swaps at shorter horizons, much like the signal from market-based overnight index swap (OIS) rates, could be more informative than previously thought.

Building on this idea, we take another look at forecast performance with a more recent sample from 2004 to the present day. For the surveys, we use the Blue Chip Economic Indicators monthly

¹For a broad overview of the limitations in surveys, see Chapter 13.5 of [Diercks and Jendoubi \(2022\)](#).

²See [Diercks, Tanaka and Cordova \(2021\)](#) for additional evidence as it relates to interest rate expectations.

³There is supporting evidence of the notion that risk premiums are smaller at the short-end from models such as [D’Amico, Kim and Wei \(2018\)](#). Their 1-year-ahead inflation risk premium is on average -1 basis point (with an absolute value of 12 basis points), while the inflation risk premium at the 5-year horizon is on average 22 basis points (with an absolute value of 31 basis points) based on data going back to the early 1980s. Moreover, standard theoretical models typically imply inflation risk premiums that increase in absolute value with the horizon when shocks to inflation are persistent as in the data. For instance, [Kliem and Meyer-Gohde \(2022\)](#) also finds that the inflation risk premium is considerably smaller at shorter horizons, about 10 basis points, on average, versus 60 basis points at the 10-year horizon.

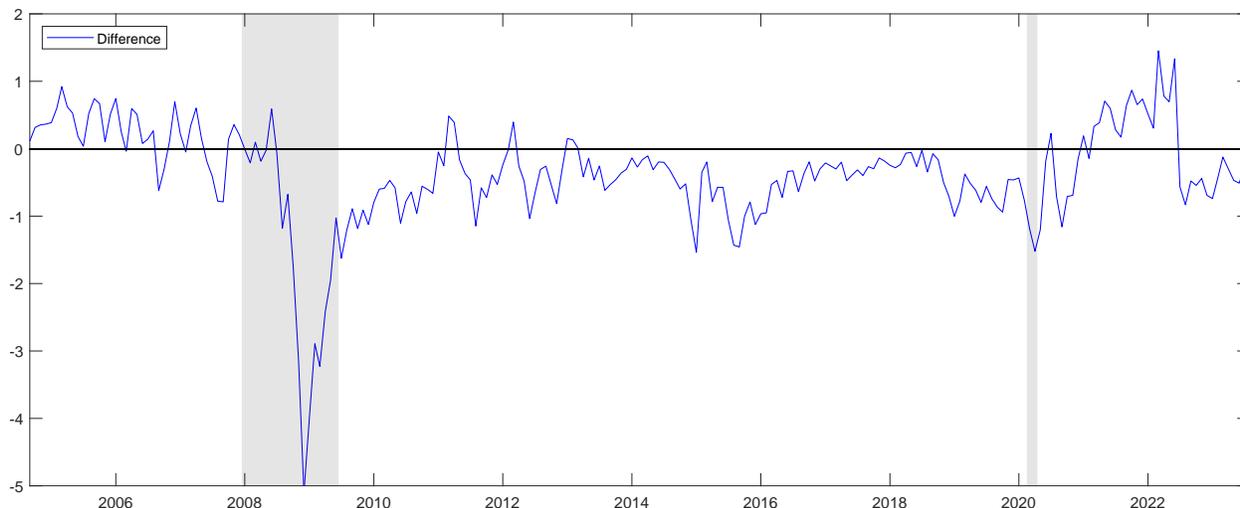


Figure 2: **Monthly Difference Between Inflation Swaps and Surveys**

Note: This figure shows the one-year inflation swap rate minus the Blue Chip Economic Indicators Forecast of one-year expected inflation. Gray bars denote NBER recessions. Source: Blue Chip; Barclays, Bloomberg.

consensus expectation of CPI inflation that extends out beyond a year. For the inflation swaps, we use measures derived from their respective financial market quotes. We start the sample in 2004 because this is when we first have data for the inflation swaps. We evaluate a number of measures of efficacy, moving beyond the more commonly used measures of mean absolute error (MAE) or mean squared error (MSE). For instance, we ask which measure is most often the best forecast? Which forecast is more likely to be better when inflation is above target? Which measure tends to outperform when there is a meaningful gap between the swaps and the surveys? Conducting these exercises is useful for distinguishing the two measures of expectations because their limitations over time are likely to reduce forecast performance. For instance, if risk premiums are consistently contaminating inflation swaps, we would expect this to reduce their forecast performance. Alternatively, if rigidities and under-reaction to incoming news is affecting surveys, we would expect this to be associated with lower forecast performance when conditions are rapidly shifting.

On the surface, our findings may at first appear somewhat consistent with previous studies documenting the strength of surveys. In terms of mean squared error, the surveys outperform the swaps over the full sample. However, this finding appears to be fundamentally related to the poor performance of inflation swaps during several months of the Global Financial Crisis (GFC) and the early months of the pandemic.⁴ When we exclude these relatively few periods when inflation swaps projected deflation, the inflation swaps outperform the surveys. Moreover, we find that the inflation swaps come closer to forecasting realized inflation about 56% of the time (and is even higher for the last ten years).

We also find a number of conditions (that are relevant for today) in which swaps tend to

⁴There are several studies documenting the severe dislocation of the TIPS market during the GFC as Lehman Brothers was forced to liquidate a vast number of these securities as their collateral had to be sold at firesale prices. See [Campbell, Shiller and Viceira \(2009\)](#), [Hu and Worah \(2009\)](#) and [Faust and Wright \(2013\)](#) for additional discussion.

outperform the surveys. First, we find that when inflation is above its historical median of 2 percent over the full sample, the inflation swaps outperform the surveys across all of the criteria we consider. Likewise, the inflation swaps more frequently come closest to realized inflation during these time periods. In addition, we find that when the gap (inflation swaps minus surveys) is more positive than -30 basis points (the median of the distribution for the gap), the inflation swaps again outperform the surveys. These results suggest that potentially greater weight should be placed on inflation swaps more recently.

In light of these findings, we look into whether placing weight on both surveys and inflation swaps can lead to superior forecasts. We find that the weights that minimize the mean absolute error are roughly half on swaps and half on surveys. This finding that forecast combinations lead to superior results is consistent with a large literature citing the benefits of such diversification (see [Diercks and Carl \(2019\)](#) for a recent application to the federal funds rate and [Fulton and Hubrich \(2021\)](#) with respect to inflation). While this weighting scheme is based on the full sample, if we were to instead use rolling ten year windows, we find that the optimal weight on inflation swaps has been steadily increasing to 0.7 most recently.

Moving towards dynamic weights and to capitalize on the relative advantage of each forecast, we consider a combination of forecasts in a smooth transition framework (as in [Chan and Tong \(1986\)](#) and [Teräsvirta \(1994\)](#)). This smooth transition model allows for time-varying weights to depend upon a transition variable that might indicate when one forecast performs substantially better than another. For our purposes, we allow the level of the inflation swap to determine the relative weight the model places on the survey versus the swap. Consistent with the poor performance of the inflation swaps during episodes of illiquidity, the framework endogenously places nearly all its weight on surveys when the inflation swap is projecting deflation and puts most of its weight on the swap when it is projecting above the threshold of 1.0 percent. We find this particular forecast, which we refer to as the DCS^2 measure of inflation expectations, has statistically significant superior predictive ability across a wide range of conditioning variables, consistently dominating the raw inflation swaps and surveys in addition to the equal weights measure.

Of course, a number of caveats are in order for any type of analysis such as this. One could easily argue that our sample is relatively short, which we would agree with, but also note that it is generally larger than most of the studies we cite. Second, given the relatively short sample, it's possible that one measure may outperform another measure based on luck and that these measures could switch places over time, as has frequently happened in the past ([Atkeson and Ohanian \(2001\)](#), [Fisher, Liu and Zhou \(2002\)](#), [Stock and Watson \(2007\)](#) and [Stock and Watson \(2010\)](#) all discuss instability in relative forecasting performance over time). Third, one might question our exclusive focus on inflation swaps and survey-based expectations when other variables might provide superior forecasts. We prioritize these measures due to their intrinsic reflection of actual expectations. Using alternative indicators lacking a direct connection to genuine inflation expectations may lead to contradictions. For instance, a GDP-based inflation forecast could rise with new data, but if swaps or surveys remain unchanged, it becomes difficult to argue that this

accurately reflects shifts in inflation expectations. And lastly, one could argue that none of these expectations are great forecasts and the differences between them may be small. Nonetheless, we believe there is inherent value in trying to determine which measures of expectations are giving the cleanest signals, and evaluating their relative forecast performance is just one aspect that can help us move in that direction.

The remaining structure of the paper is as follows. Section 1.1 covers the related literature more extensively. Section 2 describes the data we use for our forecasting exercises. Section 3 provides the forecasting results and Section 3.1 details the conditional results. Section 4 discusses the optimal weighting of the inflation swaps versus the surveys, and Section 5 concludes.

1.1 Related literature.

Our paper contributes to an extensive literature on inflation expectations and inflation forecasting with the use of surveys and market-based measures. For surveys, we focus our attention on the views of professional forecasters: for a recent overview of the literature on household inflation expectations, see [D’Acunto, Malmendier and Weber \(2022\)](#). We focus on professionals because a common finding in the literature is that they tend to outperform households in terms of forecasting. For example, [Thomas \(1999\)](#), [Ang, Bekaert and Wei \(2007\)](#), [Trehan \(2015\)](#), [Berge \(2018\)](#), [Aruoba \(2020\)](#), and [Bennett and Owyang \(2022\)](#) all generally find that professional forecasters outperform household expectations such as the Michigan Survey of Consumers.⁵

[Ang, Bekaert and Wei \(2007\)](#) along with [Faust and Wright \(2013\)](#) provide some of the most extensive comparisons between model-based forecasts and survey-based forecasts and conclude that surveys are hard to beat. In contrast to our work, direct market-based measures are not included in these studies either because they did not yet exist or the sample was too short. More recent studies such as [Bauer and McCarthy \(2015\)](#) and [Grothe and Meyler \(2018\)](#) do compare swaps to surveys and use the Survey of Professional Forecasters to find that surveys continue to outperform.⁶ In contrast to these studies, our sample extends to 2022 and uses monthly data available from the Blue Chip Economic Indicators, effectively providing us with five to seven times the number of observations. [Bennett and Owyang \(2022\)](#) also evaluate the recent performance of various inflation forecasts but exclude swaps at the one-year horizon. [Martinez \(2020\)](#) also compares survey and market-based measures and finds their combination can lead to improved forecasts but focuses on long-term expectations. Lastly, [Verbrugge and Zaman \(2021\)](#) is closest to our work and also evaluates one-year forecasts of inflation. They evaluate a wider set of measures but use a one-year inflation swap rate that is extracted from the model of [Haubrich, Pennacchi and Ritchken \(2012\)](#). In contrast, we use a direct reading of the inflation swap for our analysis that does not attempt to control for risk premiums and we properly control for the indexation lag.

On the surface, not attempting to control for risk premiums would seem ill-advised given that

⁵One exception is [Coibion and Gorodnichenko \(2015\)](#), who finds the Michigan Survey of Consumers outperform the one-year ahead expectations of the SPF when it comes to the missing disinflation from 2009 to 2011.

⁶[Grothe and Meyler \(2018\)](#) also evaluate Euro area inflation and technically do not directly compare the surveys and the inflation swaps due to the indexation lag, which we will further discuss in section 2.1.

they have been extensively studied and documented for market-based measures of inflation expectations. There are numerous studies that find that inflation risk premiums have been an important component and that much of the movements in longer-dated yields can reflect risk or liquidity premiums in addition to expectations. For instance, [D’Amico, Kim and Wei \(2018\)](#) argue that TIPS relatively lower liquidity makes it difficult to take signal from the implied inflation expectations due to liquidity premiums. [Christensen, Lopez and Rudebusch \(2010\)](#) also determines an important role for inflation risk premiums but finds that on average they are closer to zero.⁷ [Grishchenko and Huang \(2013\)](#) account for the indexation lag in TIPS and also document average inflation risk premiums of about 15 basis points at the ten-year horizon. [Chernov and Mueller \(2012\)](#) evaluate a longer sample while incorporating surveys and estimate inflation risk premiums to be close to 2 percent at a similar horizon. [Abrahams, Adrian, Crump, Moench and Yu \(2016\)](#) use an observable factor to adjust TIPS spreads for their liquidity (while not relying on surveys) and estimates an average inflation risk premium of about 50 basis points at the 5- to 10-year horizon.⁸ However, all of these studies focus on longer-horizon expectations and in fact, nearly all of them exclude horizons below two years, as also pointed out by [Aruoba \(2020\)](#). Thus, the literature’s implications for premiums at the short-end of the curve (e.g., one year horizon), which is the focus of our study, is not as well documented and its not obvious these premiums at shorter horizon should be as consequential.

On that note, there’s a related literature documenting the lack of sizable risk premiums at shorter horizons for interest rate expectations. [Longstaff \(2000\)](#) focuses on repo rates and finds no evidence of risk premiums for maturities up to three months, while [Della Corte, Sarno and Thornton \(2008\)](#) extend the sample and use more powerful tests to confirm that risk premiums in very short-term repo rates are too small to be economically important. [Downing and Oliner \(2007\)](#) explore the commercial paper market with horizons up to 90 days and find evidence in support of the generalized expectations hypothesis. Likewise, [Hamilton \(2009\)](#) looks at daily changes in fed funds futures and determines contracts up to 2.5 months out exhibit risk premiums that are of little economic importance. [Sack \(2004\)](#) and [Durham \(2003\)](#) similarly find modest time-variation in risk premiums for fed funds futures at shorter horizons. In terms of overnight index swap rates, [Lloyd \(2020\)](#) finds that excess returns on OIS rates at the daily frequency are insignificantly different from zero out to two years and are insignificant out to three years when controlling for the Global Financial Crisis. Likewise, [Wang and Yang \(2018\)](#) find that average excess OIS returns are insignificantly different from zero over their full sample. There are also several studies emphasizing that risk premiums in treasury and money market derivatives based on predictive regressions (which tend to be sizable) more likely reflect nonzero expectational errors rather than risk compensation (for instance, see [Froot \(1989\)](#), [Ferrero and Nobili \(2009\)](#), [Bacchetta, Mertens and Van Wincoop \(2009\)](#), [Ichiue and Yuyama \(2009\)](#), [Chun and Chun \(2013\)](#), [Cieslak \(2018\)](#), [Diercks and Carl \(2019\)](#)

⁷In their sample from 2003 to 2008, the average 5- and 10-year inflation risk premiums were both about -5 basis points, while the [D’Amico, Kim and Wei \(2018\)](#) averages were 36 and 64 basis points, respectively.

⁸[Chen, Liu and Cheng \(2010\)](#) summarizes some empirical estimates of the inflation risk premium in the literature that are broadly consistent with the above mentioned studies.

and [Schmeling, Schrimpf and Steffensen \(2020\)](#)). Thus, if money market derivatives are found to exhibit little to no risk premiums at shorter horizons, derivatives such as inflation swaps at shorter horizons may have similar risk characteristics.⁹

Inflation swaps, which are the focus of our study, have also frequently been utilized in place of TIPS. [Haubrich, Pennacchi and Ritchken \(2012\)](#) combine inflation swaps and survey forecasts for their estimation and judged that inflation swaps face fewer distortions compared to TIPS yields.¹⁰ [Gospodinov and Wei \(2016\)](#) also incorporate inflation swaps (in addition to oil future prices) and find this leads to increased forecasting performance. [Fleming and Sporn \(2013b\)](#), [Fleming and Sporn \(2013a\)](#) study the market for inflation swaps and conclude they are reasonably liquid with modest bid-ask spreads.

2 Data

In this section, we detail the two main sources that we use for our analysis: inflation swaps and survey-based expectations. Our focus for these measures is a one-year inflation rate.

2.1 Inflation Swaps

The inflation swaps are derivatives traded in a dealer-based over-the-counter market. To enter such a contract, the buyer pays a fixed fee and in return will receive payment equal to the realized ("floating") inflation rate for a particular time period, with net cash flows exchanged at contract maturity. The fixed fee is said to be the expected rate of inflation over the contract horizons, but may also reflect a small insurance, or "risk" premium. We use quotes provided by Barclays and Bloomberg to construct monthly data on inflation swap rates. The data covers the period from 2004 to the present and we use the 8th day of each month to be consistent with the surveys.

The market for inflation swaps is generally quite liquid as discussed in [Fleming and Sporn \(2013b\)](#) and [Fleming and Sporn \(2013a\)](#). While TIPS-based markets suffer from a number of liquidity issues (as highlighted by [To and Tran \(2019\)](#), [Fleckenstein, Longstaff and Lustig \(2014\)](#), [Campbell, Shiller and Viceira \(2009\)](#), [Pflueger and Viceira \(2011a,b\)](#)), these issues are less serious for inflation swaps. One key differentiating factor is that only the net cash flows are exchanged at the maturity of the swap contract, while transactions in the TIPS market require the whole notional value to be exchanged. To be clear, this does not mean that liquidity issues in the TIPS market can not bleed into the inflation swaps, and inflation swaps can be more subject to counterparty credit risks. Nonetheless, several studies have found that swaps, especially at the shorter horizons, can be of considerable value in terms of gauging shorter-horizon expectations relative to similar

⁹For instance, the volatility of the federal funds rate is about 1.6% since 2004, while the volatility of CPI 12-month inflation is 1.9%. Moreover, studies such as [D'Amico, Kim and Wei \(2018\)](#) and [Kliem and Meyer-Gohde \(2022\)](#) estimate inflation risk premiums at the one-year horizon to be less than 10 basis points on average, which is economically small and roughly consistent with evidence from money market derivatives.

¹⁰[To and Tran \(2019\)](#), [Fleckenstein, Longstaff and Lustig \(2014\)](#), [Campbell, Shiller and Viceira \(2009\)](#), [Pflueger and Viceira \(2011a,b\)](#) are also important studies documenting various liquidity issues in TIPS markets.

maturity TIPS (for example, see [Haubrich, Pennacchi and Ritchken \(2012\)](#) and [Gospodinov and Wei \(2016\)](#)).¹¹

It must also be noted that inflation swaps have an indexation lag that is similar to TIPS. The indexation lag is 3 months and implies that the realized inflation that is relevant for the contract is based on the price level three months before the start date and the price level nine months after that date, thus gauging the growth rate in the price level from $t - 3$ to $t + 9$. This timing is used because the CPI is only known with a lag.

Figure 3 provides a diagram documenting the timing of CPI releases, illustrating why the one-year inflation swap at the time of initiation covers 11 unknown months of inflation, plus one known month. The key to understanding why there are 11 months, and not 9 months, of unknown inflation is to recognize that CPI releases for a given month are not known until the middle of the following month. Thus, on January 8, 2023, the latest available CPI level corresponds to November 2022, which was released on December 13, 2022. The second key is to note that an inflation rate requires two price levels in order to be computed. Thus, while the $t-3$ (October) and $t-2$ (November) price levels are known at time t , the $t-1$ (December) and time t (January) inflation rates are still unknown on January 8, 2023. The historical December inflation rate, the current (January) inflation rates, along with the nine inflation rates through October 2023, make up the 11 unknown months of inflation for which we are extracting expectations.

Given that the underlying inflation swap incorporates the observed inflation rate for $t-2$ (i.e. the growth in the price level between $t-3$ and $t-2$), we can remove its influence on the inflation swap price so that our measure only reflects expectations for the 11 months of inflation that are not yet realized. Absent this adjustment, one month covered by the inflation swap would have zero forecast error, which would reduce the overall measure of forecast error. To address this issue, we compute the adjusted swap rate as follows:

$$\pi_t^{adj\,swap} = \left(\left(\left(1 + \frac{\pi_{t-3,t+9}^{swap}}{100} \right) * \frac{1}{1 + \frac{\pi_{t-3,t-2}^{realized}}{100}} \right)^{12/11} - 1 \right) * 100 \quad (1)$$

The inflation swaps payout is linked to the non-seasonally adjusted CPI for all urban consumers. Given that the 1-year swap is technically a bit less than that, its possible that seasonality effects may also be an issue. However, as also emphasized by [Bauer and McCarthy \(2015\)](#), controlling for seasonality requires estimating seasonality effects, which induces its own uncertainty and potential estimation error.¹² Thus, we will use the direct adjusted reading of the inflation swap for our purposes to remain as parsimonious as possible.¹³

¹¹While TIPS may suffer from a number of liquidity issues, it should be noted that their trading volume is much higher at close to \$20 billion a day versus the \$1 billion a day for inflation swaps (based on data from 2017, see [To and Tran \(2019\)](#)).

¹²When [Bauer and McCarthy \(2015\)](#) attempted to control for seasonality, they found it did not improve the performance of the inflation swaps.

¹³In addition, tests for seasonality such as regressions on monthly indicator variables are insignificant, suggesting seasonality is not a serious issue for this time series.

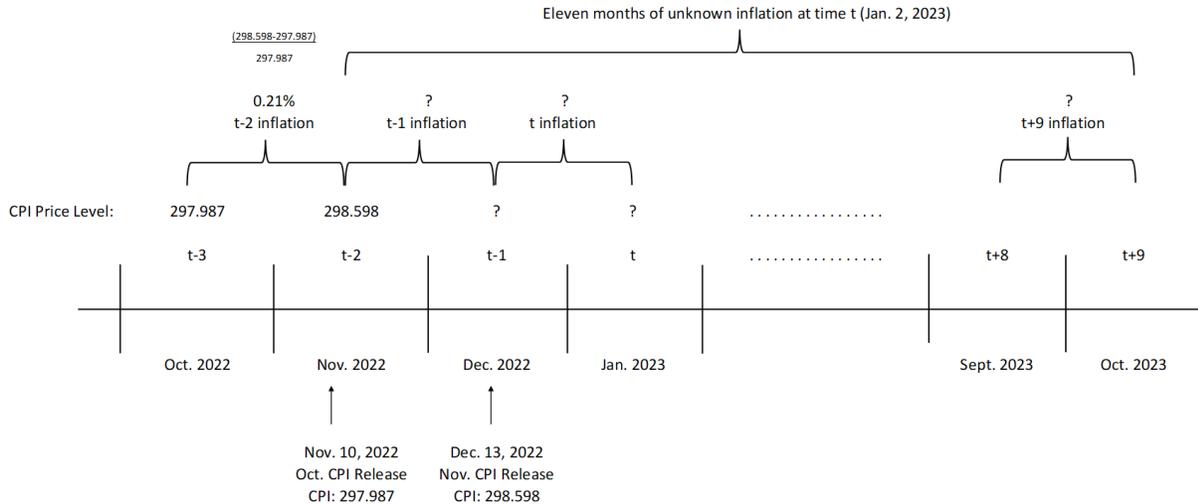


Figure 3: **Diagram of Timeline for Inflation Swap on January 8, 2023**

Note: This diagram shows the timeline of the relevant months for constructing the inflation expectations based on the inflation swap. The date January 8, 2023 corresponds to one of the observations in our sample.

2.2 Survey-based Expectations

For the surveys, we use the Blue Chip Economic Indicators monthly survey, which has been conducted since 1982 and produced by Wolters Kluwer. This differs from the Survey of Professional Forecasters, which is available quarterly. The survey responses are usually submitted three business days before publication and is typically released to the public on the tenth of each month. We use the consensus forecasts, which are the calculated cross-sectional means based on the 40 to 50 respondents.

Respondents provide a seasonally-adjusted annualized growth rate of the average price level for the current quarter and quarters up to the end of the following year. For example, the first quarter of 2023’s average inflation rate is computed by taking the average price level for the months of October, November, and December (i.e., 2022Q4) and computing the growth rate to the expected average price level for the months of January, February, and March (2023q1).

To extract the expected inflation rate for an 11-month period that is precisely comparable to the swap rate requires some additional gymnastics, which we outline in Table 1. Since the Blue Chip provides expectations for the annualized growth rate of the quarterly average price level, we will need to interpolate to infer the Blue Chip expected price level in October 2023. Again, that price level determines the payoff to a 1-year swap purchased the first week of January.

Since the price levels for the quarter ending in December have not been produced and published, we use the Blue Chip’s backcast to calculate the expected average price level for the fourth quarter (Q4) of 2022. We then iterate forward from this 2022Q4 average price level, multiplying it by each subsequent expected inflation rate up to 2023Q4 of the next year.

To pinpoint the monthly price level for October 2023—which is nine months after our starting point in January 2023—we assume that our calculated 2023Q4 average price level represents the

Time t	Qtr	Month	Monthly Realized Price Level	Average Qtrly Price Level	BC Consensus Inflation	Notes
	Q3	July	294.628			
	Q3	Aug	295.320	295.496		
	Q3	Sept	296.539			
-3	Q4	Oct	297.987			
-2	Q4	Nov	298.598	298.391	3.92%	= Q3 Avg Price Level * (1 + 3.92 / 400)
-1	Q4	Dec	?			
0	Q1	Jan	?			
1	Q1	Feb	?	300.815	3.25%	= Q4 Avg Price Level * (1 + 3.25 / 400)
2	Q1	Mar	?			
3	Q2	Apr	?			
4	Q2	May	?	303.029	2.94%	= Q1 Avg Price Level * (1 + 2.94 / 400)
5	Q2	June	?			
6	Q3	July	?			
7	Q3	Aug	?	305.050	2.67%	= Q2 Avg Price Level * (1 + 2.67 / 400)
8	Q3	Sept	?			
9	Q4	Oct	?	306.264		
	Q4	Nov	?	306.871	2.39%	= Q3 Avg Price Level * (1 + 2.3 / 400)
	Q4	Dec	?			We want 9 months ahead so need to interpolate between 306.871 and 305.050 (306.871-305.050)/3*2+ 305.050 to get Oct price level assuming Aug. is 305.050

Table 1: **Blue Chip Inflation Computations for Jan. 2023**

Note: This provides an example of how the Blue Chip Economic Indicators CPI Inflation rate is computed for Jan. 2023.

midpoint of the quarter (November). We then interpolate this value between the months of August (midpoint of 2023q3) and November 2023. Finally, we calculate our expected inflation rate by calculating the growth in the realized price level from time t-2 (Nov. 2022) through the expected price level at t+9 (Oct. 2023), both of which are in bold in Table 1, which correspond to the 11 months of unknown inflation.

2.3 Realized inflation

The object of interest in terms of forecasting is the annualized 11-month inflation rate for the CPI All Urban Consumers Seasonally Adjusted index. Specifically, we compute the rate as $100 \cdot \left(\left(\frac{CPI_{t+9}^{Level}}{CPI_{t-2}^{Level}} \right)^{\frac{12}{11}} - 1 \right)$ to be consistent with the swap. Note that technically, the Blue Chip forecasts are based on the seasonally adjusted rate while the inflation swap is tied to the non-seasonally adjusted rate. Thus, by using seasonally adjusted price levels, our results are likely biased in favor of the surveys and against the swaps. Therefore, any results in favor of the swaps should be viewed as conservative or in some sense a lower bound.

Figure 4 shows the resulting time series for our three key metrics – the inflation swaps (blue), the surveys (red), and the realized inflation rate (black). A few things are worth noting: (1) the surveys seem to stay close to 2% for most of the sample, (2) the inflation swaps are relatively more volatile, but so is the realized inflation rate, (3) the inflation swaps remained below the surveys from 2011 through 2020, and the realized inflation rate was also typically lower, and (4) neither the inflation swaps nor the surveys predicted the large increase of inflation more recently, though the inflation swaps come somewhat closer.

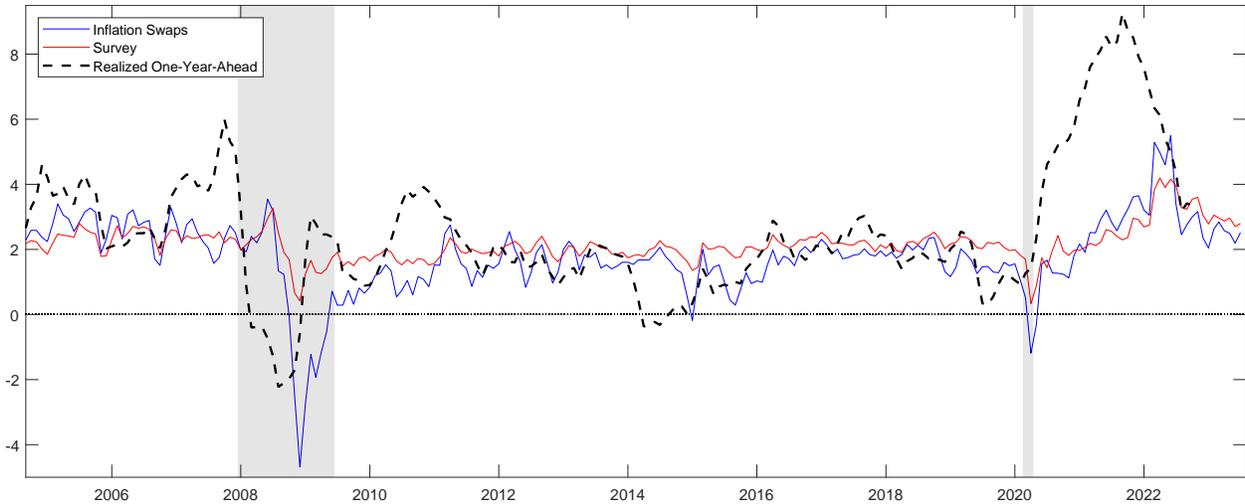


Figure 4: **Time Series of Inflation Swaps, Surveys, and Realized Inflation Rate**

Note: This figure compares the monthly inflation swap rates (blue) to surveys (red) and the realized one-year inflation rate (dashed black). NBER recessions are shown in gray bars.

2.4 Comparing Distributions

Before moving to the forecast performance, we can compare the distributions of the expectations to realized inflation. Figure 5, left panel, shows a histogram of the survey-based expectations compared to the realized inflation rates. The right panel shows the associated inflation swap rates compared to realized inflation. One can see that, relative to the observed outcomes for realized inflation, the surveys tend to put a disproportionate number of forecasts bunched close to 2%. In contrast, the inflation swap rates are generally more dispersed.

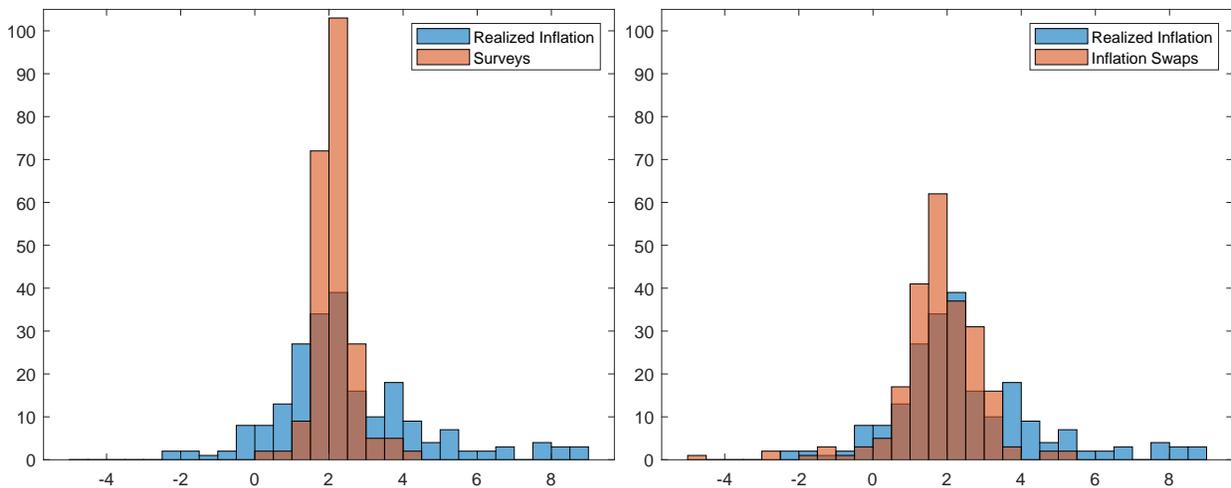


Figure 5: **Histogram of Forecasts vs Realized: Surveys (Left) and Swaps (Right)**

Note: The horizontal axis reflects the forecast or realized inflation in percentage points, while the vertical axis captures the frequency.

Table 2 shows summary statistics consistent with the previous figure. The mean for the inflation

swaps is lower than the surveys and also the realized inflation rate, while the survey mean is closer to the mean of the realized inflation rate. In contrast, the variance of survey expectations is considerably lower than the realized inflation rate. The variance of the inflation swaps is notably higher than for survey expectations but still much lower than the variance of the realized. One potential takeaway from this exercise is that while the surveys appear less biased over the full sample, the extremely low variance may be consistent with some rigidities that reduce the forecast performance of the surveys.

One-Year Expectation	Full Sample	
	Mean	Variance
Inflation Swaps	1.76	1.36
Blue Chip Survey	2.14	0.26
Realized Inflation Rate	2.54	4.28

Table 2: **Mean and Variance of Expectations and Realized Inflation**

Note: Full sample is from September 2004 to October 2022 (one-year realized inflation did not exist beyond this date at the time of writing).

3 Forecasting Results

This section documents the forecast performance of inflation expectations coming from the Blue Chip survey and inflation swaps. Table 3 shows the forecast performance in terms of the mean squared error across several samples. Specifically, we compare expectations using the full sample (2004 to present) and two sub-samples where we exclude the few time periods in which the inflation swaps projected deflation and were widely thought to be hindered by severe liquidity issues. Specifically, applying the “No Deflation” criterion excludes 8 months during the Global Financial Crisis (October 2008 to May 2009), 3 months during the beginning of the pandemic (March 2020 through May 2020) and, lastly, January 2015, when a short episode of illiquidity was associated with the removal of the Swiss franc peg to the euro.¹⁴ Thus, the no-deflation criterion represents 12 out of the 218 months, or just about 5% of the sample.

Our first main finding is that the Blue Chip survey slightly outperforms the inflation swaps over the full sample. This finding seems consistent with many of the previous studies that suggest surveys seem to dominate in terms of forecasting (see [Ang, Bekaert and Wei \(2007\)](#), [Gil-Alana, Moreno and Pérez de Gracia \(2012\)](#), [Faust and Wright \(2013\)](#), and [Bauer and McCarthy \(2015\)](#)). An alternative measure of performance involves counting the percentage of months over the sample that the inflation swaps ended up closer to the realized inflation rate than the Blue Chip survey. Here, we find that on 56.0% of the months in our sample, the inflation swaps outperform the

¹⁴Our results are little changed if we do not exclude January 15 for the “No Deflation” sample.

surveys. In other words, despite the under-performance based on mean squared error, and despite the common interpretation of survey point forecasts represent modal forecasts, the inflation swaps were closer to the realized inflation rate more often.

One-Year Expectation	Full Sample	No Deflation	
		Full Sample	Last 10 Years
Panel A: Mean Squared Error			
Inflation Swaps	4.03	3.68	3.79
Blue Chip Survey	3.93	3.99	4.32
Panel B: Mean Absolute Error			
Inflation Swaps	1.38	1.29	1.21
Blue Chip Survey	1.35	1.34	1.30
% of Months Swaps Beat Surveys	56.0%	57.5%	59.7%
Observations	218	207	139

Table 3: Mean Squared Errors and Mean Absolute Errors

Note: Full sample is from September 2004 to October 2022 (one-year realized inflation did not exist beyond this date at the time of writing). The No Deflation sample omits the 12 months out of 218 in which the inflation swaps projected deflation. The Last 10 years starts in January 2011 and omits 4 months out of 134 in which inflation swaps projected deflation.

Nonetheless, it is well-known and documented by other studies such as [Campbell, Shiller and Viceira \(2009\)](#), [Hu and Worah \(2009\)](#) and [Faust and Wright \(2013\)](#) that financial market-based inflation forecasts during the Global Financial Crisis (GFC) were severely affected by the forced liquidation of Lehman Brothers. To better understand the relative influence of this episode as well as the onset of the pandemic, we exclude these time periods and show the results in the second column of Table 3. Without the deflationary periods, we see that the inflation swaps now emerge more clearly as the measure with the lowest mean squared error, with an MSE that is about 8% lower than the surveys. We also see that the proportion of observations in which the swaps outperform the surveys rises to 57.5%.

We see a similar dynamic if we focus on the non-deflationary periods over the last ten years, as shown in column 3 of Table 3. Again, the swaps outperform the surveys, with inflation swaps being 12% lower. These results indicate that, outside of a few months during the global financial crisis and early months of the pandemic, the inflation swaps have been a relatively better forecaster than the surveys; moreover, even if we do not exclude these periods, the inflation swaps are more likely to come closest to the realized inflation rate.

Table 3 Panel B provides a similar message based on the mean absolute error criterion. The mean absolute error criterion is also useful because it is relatively more robust in that it does not

A. Full Sample

1Y Exp.	MSE	USPA	CSPA						
			INFL	UR	Δ IP	Δ C	Δ RPI	VIX	FFR
Infl. Swaps	4.03	0.399	0.005***	0.001***	0.001***	0.005***	0.259	0.048**	0.135
BC Survey	3.93	1.000	0.001***	0.002***	0.004***	0.001***	0.569	0.019**	0.040**

B. No Deflation Sample

1Y Exp.	MSE	USPA	CSPA						
			INFL	UR	Δ IP	Δ C	Δ RPI	VIX	FFR
Infl. Swaps	3.68	1.000	0.175**	0.077*	0.309**	0.437**	0.613**	0.520**	0.890**
BC Survey	3.99	0.165	0.001***	0.001***	0.003***	0.001***	0.007***	0.001***	0.004***

Table 4: **Significance Tests for Forecasting Ability of Swaps vs Survey**

Note: The first column shows the mean squared error for the inflation swaps and survey forecasts. The second column shows the p-value for the test for unconditional superior predictive ability (USPA), while the remaining columns show the p-values when testing for conditional superior predictive ability (CSPA). The stars, *, **, ***, denote 90%, 95%, and 99% significance and indicate rejection of the null that a given forecast weakly dominates the alternative forecasts (and, in the case of CSPA, across all states for a given conditioning variable). INFL refers to annual CPI inflation, UR is the unemployment rate, Δ IP is annual growth in industrial production, Δ C is annual growth in personal consumption expenditures, Δ RPI is annual growth in real personal income, VIX is a market-based measure of uncertainty, and FFR is the Federal Funds Rate.

place extra weight on outliers, in contrast to the mean squared error. The first column of Table 3 Panel B shows once again that the surveys outperform the inflation swaps. However, when we move to the non-deflationary time periods we again see that the inflation swaps outperform the surveys.

To get a sense of the statistical significance of the difference between the two forecasts across the samples, we rely on the test for conditional superior predictive ability (CSPA) from Li et al. (2021). CSPA tests whether a given forecast method is weakly superior to a set of alternative methods across all states of a conditioning variable. More formally, the null hypothesis is that the benchmark forecast’s conditional expected loss is no larger than another forecast uniformly across all conditioning states. In other words, the best forecasts will fail to reject the null hypothesis because the competing forecasts will not statistically dominate it based on the conditioning states.

Similar to the test introduced by Giacomini and White (2006), the CSPA test captures the fact that the relative performance of two forecasting methods may differ across states of the world. This feature contrasts with unconditional tests, such as Diebold and Mariano (2002), White (2000), and Hansen (2005), that only capture if a forecast method performs better than alternative method(s) *unconditionally on average*. Moreover, Li et al. (2021) shows that averaging across states can

lead to situations in which the performance of various forecasts are very difficult to statistically distinguish. By relying on an unconditional average, this approach will often integrate out periods or states in which there is a clear difference in the relative performance of different forecasts.

Table 4 compares the forecasting ability for the swaps and survey in both the full sample (top panel) and the subsample where the swaps do not forecast deflation (bottom panel). The first column shows the mean square error previously mentioned, which the survey outperforms the swaps in the full sample but the swap is better in the no-deflation subsample. The second column shows the p-value from a test of unconditional superior predictive ability (USPA) as in Hansen (2005). Similar to the CSPA, the null hypothesis of the USPA is that a given forecast performs at least as well as the other, alternative forecast. For both the swaps and the survey, we cannot reject this null for the USPA in either the full or no-deflation subsample. Therefore, we do not find that either one of the forecasts significantly outperforms the other on average.

The last six columns show the p-values from a CSPA test for an array of different conditioning variables: (i) annual CPI inflation, (ii) the unemployment rate, (iii) industrial production growth, (iv) consumption growth, (v) employment growth, and (vi) the VIX. These variables were chosen due to their macroeconomic importance and breadth, thereby providing robustness for our results. For the full sample, we reject the null that the swaps perform at least as well as the survey for all six conditioning variables. We also reject the null for the surveys. In other words, for the full sample, both forecasts have conditioning states in which they are significantly dominated by the other forecast and thus both reject the null.

To provide some more intuition for how the CSPA test works in this context, it helps to plot the expected loss differential against the conditioning variables. In Figure 6, we show the expected loss for the realized inflation conditioning variable (left panel) and the unemployment rate (right panel). In this example, $\Delta Loss$ reflects the expected squared loss for the survey minus the expected squared loss for the swap. Thus, if the $\Delta Loss$ is significantly below zero, this would indicate the survey's expected square loss is significantly lower than the swap's at that particular state. Likewise, if its positive, it means the survey's squared forecast error is above the swaps. Only one line is plotted for the 95% confidence band because the CSPA test is for weak dominance.

One can see from Figure 6 that for lower values of realized inflation, the survey's expected squared error is significantly less than the swap's expected squared error. In contrast, for values above 0.02 or 2%, the survey's loss is larger than the swaps. Note also that the error band is much more narrow around 2% as there are more observations in this neighborhood.

We can get a similar sense of the relative performance when the unemployment rate is the conditioning variable (right panel). For relatively low unemployment rates (which makes up most of the sample), the survey's expected squared errors are greater than the swaps. However, for periods of extremely high unemployment such as the GFC and pandemic, the surveys squared errors are significant lower.

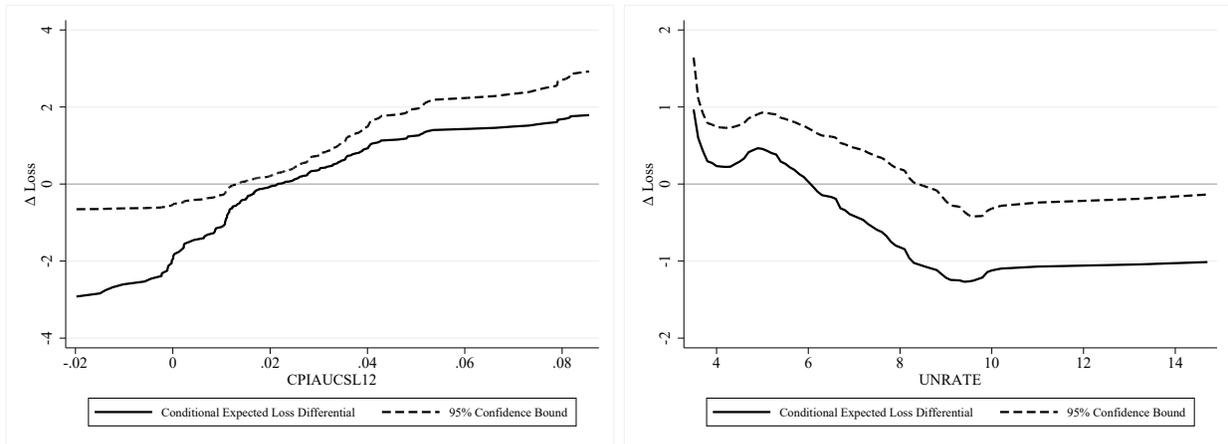


Figure 6: **CSPA Loss Differentials for Survey minus Swaps: Inflation (left), Unemployment (right)**

Note: This figure shows the expected loss differential (survey squared error minus swap squared error) across the conditioning variables of inflation and the unemployment rate.

Table 4 Panel B shows the swaps strike back when using the No Deflation sample. We fail to reject the null that the swaps weakly dominate for almost all seven conditioning variables, whereas we do reject the null for each of the seven cases when using the survey as the benchmark. This finding provides evidence that the swaps outperform the survey in a statistically significant manner when episodes of illiquidity are removed.

We can again see how this works visually by plotting the conditional expected loss differentials for the No Deflation sample in Figure 7. In contrast to before, the conditional expected loss differential (surveys minus swaps) is now insignificantly negative for low inflation. Likewise, the squared errors for surveys are insignificantly lower than the swaps for high unemployment at the 95% level. Thus, when we exclude the relatively few periods in which swaps project deflation, we can see that the swaps weakly dominate the surveys across a number of conditioning variables.¹⁵

¹⁵The rest of the conditioning variables are plotted in the Appendix.

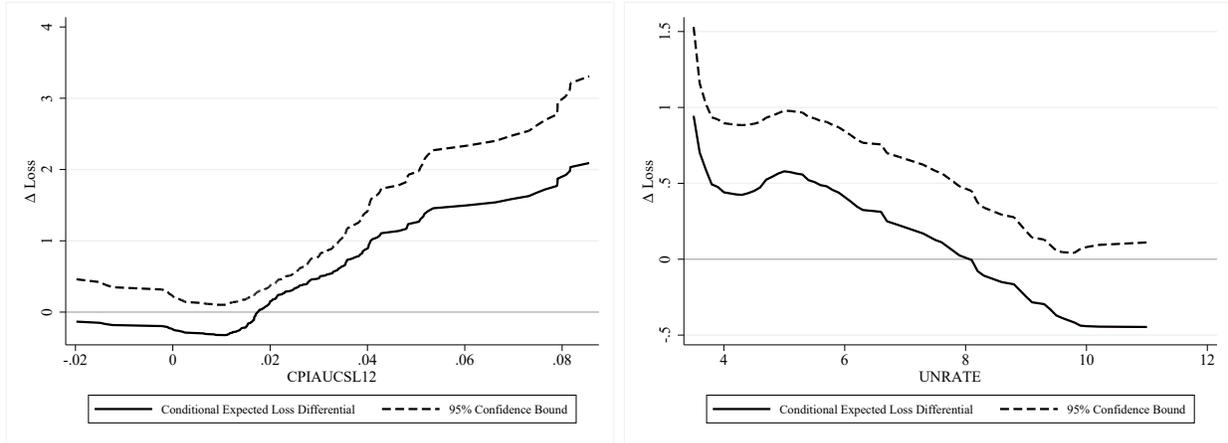


Figure 7: CSPA Loss Differentials for Survey minus Swaps (No Deflation): Inflation (left), Unemployment (right)

Note: This figure shows the expected loss differential (survey squared error minus swap squared error) across the conditioning variables of inflation and the unemployment rate for the No Deflation sample.

3.1 Conditional Forecasting Results

We can move beyond unconditional criteria and ask the following question: does one particular measure tend to outperform when realized inflation is above its historical median? Table 5 shows the relative forecasting performance for the 50% of the sample in which inflation is above its median since 2004 (the median turns out to be close to 2%). We can see that during these time periods, the inflation swaps outperform for both mean squared error and mean absolute error. Moreover, the inflation swaps come closer to the realized inflation rate when compared to the surveys 68% of the time. This is obviously relevant for the current period with inflation well above its historical median dating back to 2004.

We also note that the gap between the survey and the inflation swaps has been about -30 basis points (with the swaps being lower) over the past few months. We can also condition on this particular situation and ask what the relative performance is when the swaps minus survey is more positive than -30 basis points. Table 5 shows that this situation is also one in which the inflation swaps tend to outperform.¹⁶

4 Combining Forecasts

4.1 Time-varying optimal weights

There is a large literature showing how combining forecasts can lead to superior results through the benefits of diversification, including recent applications to forecasting the federal funds rate [Diercks](#)

¹⁶We can also compare performance strictly over the pandemic, or since March 2020, which is not shown in the table. For this period, the swaps do the best in terms of mean squared error and mean absolute error. That said, none of the measures perform well over this time period.

One-Year Expectation	Mean Squared Errors		Mean Absolute Errors	
	Inflation > Median	Swaps - Survey > -0.3	Inflation > Median	Swaps - Survey > -0.3
Inflation Swaps	4.66	5.15	1.44	1.57
Blue Chip Survey	5.71	6.06	1.63	1.74
% of Months Swaps Beat Surveys	68.1%	68.0%	68.1%	68.0%
Observations	108	108	108	108

Table 5: **Conditional Comparisons**

Note: Full sample is from September 2004 to October 2022 (one-year realized inflation did not exist beyond this date at the time of writing). The median of inflation over this sample is 2.04% and the median gap between the swaps and survey is -0.3.

and Carl (2019) and inflation [Fulton and Hubrich \(2021\)](#).¹⁷ In this section, we allow the data to inform us regarding the optimal time-varying weight on inflation swaps versus surveys. For the first exercise, we determine the weight between zero and one that is to be placed on the swaps versus the surveys using an expanding window of historical observations. The optimal weight minimizes the mean absolute error starting in 2004 and is recalculated after extending forward the sample one period at a time.

Figure 8 (left panel) shows the optimal swap weight over time for such an exercise. We can see that for the full sample (dashed blue line), the optimal weight on swaps dramatically falls close to zero during and after the GFC, consistent with its poor performance during that time period. We then see it rise to a little under half by the end of the sample. In contrast, for the No Deflation sample (red line), the optimal weight is higher and by 2022 puts more than 70% on the swap.

¹⁷For an extensive overview of model averaging and forecast combinations, see [Elliott and Timmermann \(2013\)](#).

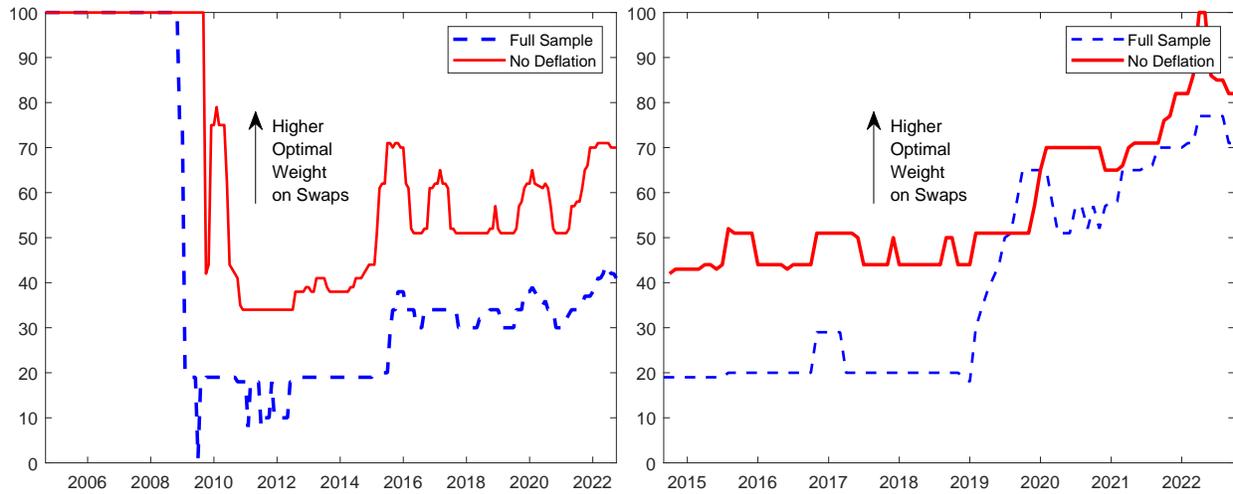


Figure 8: **Optimal Swap vs Survey Weights: Minimum Absolute Error, Expanding Window (left) and Rolling Window (right)**

Note: A weight of 100 indicates all weight being optimally placed on inflation swaps. A weight of 0 indicates all weight being optimally placed on surveys. Source: Blue Chip; Barclays, Bloomberg.

We could instead use a rolling window as opposed to an expanding window. For this exercise, we determine the optimal weight on swaps versus surveys based on the most recent ten years of forecast performance. Again, we see in the right panel of Figure ?? that, earlier in the sample, a relatively low weight was placed on the inflation swaps. As time moves forward and the GFC is dropped from the sample, the optimal weight on the inflation swaps moves to about 70%. In contrast, if the deflation period is barred from entering the sample, the optimal weight remains close to 50% for most of the sample and rises more recently to 80%.

In the right panel of Figure 9, we conduct a similar rolling 10-year window using mean squared error as the criterion. Again, with the fairly poor performance of swaps during the GFC and with this particular criterion placing greater emphasis on the errors over this time period, the inflation swaps receive less weight for longer. Once, the GFC rolls out of the 10-year sample, the weight placed on the swaps rise dramatically, moving above 70% most recently. When we exclude deflationary periods, the weight is higher and jumps to 100% by the end of the sample. The stark rise is likely due to the relatively higher forecasts coming from inflation swaps more recently.

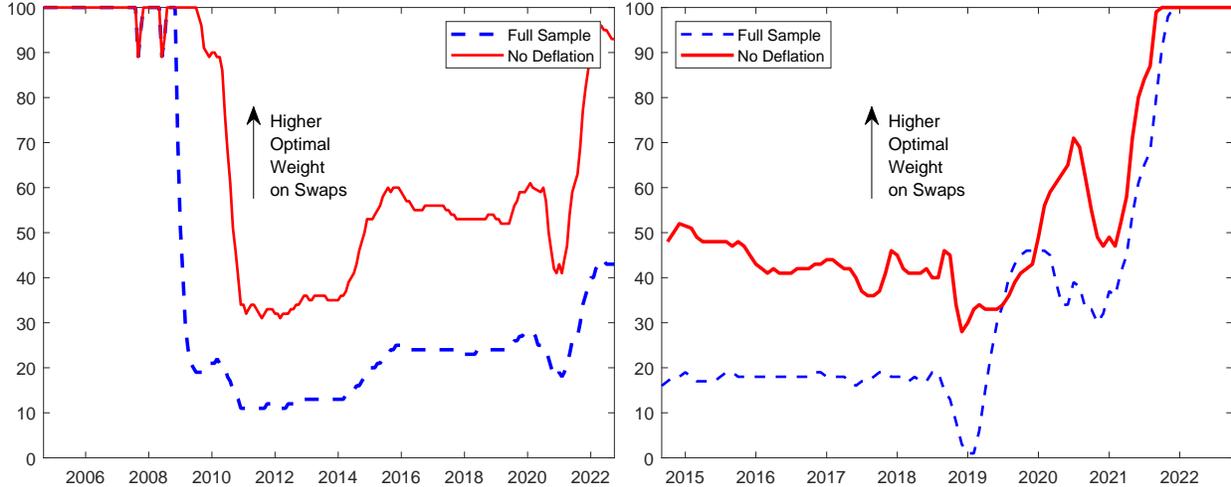


Figure 9: **Optimal Inflation Swaps Weights: Minimum Squared Error, Expanding Window (left) and 10-Year Rolling Window (right)**

Note: A weight of 100 indicates all weight being optimally placed on inflation swaps. A weight of 0 indicates all weight being optimally placed on surveys. Source: Blue Chip; Barclays, Bloomberg.

Lastly, in the left panel of Figure 9, we conduct the exercise for the expanding window for the mean squared error. Consistent with the other exercises, when the GFC is included in the sample (blue line) the optimal weight falls dramatically in 2008. With the errors being squared over this time period, the additional emphasis on the large errors over this time period puts downward pressure on the optimal weight placed on the inflation swaps for the rest of the sample. By the end, the optimal weight approaches about 45%. In contrast, when the deflationary periods are excluded, we see a similar decline but the optimal weight rises to a greater extent in the years that follow, with the optimal weight topping 90% by 2022.

4.2 Smooth Transition Regime-Switching Model

The previous section provides two key results: (i) inflation swaps outperform the survey during “normal” times, and (ii) the survey outperforms during certain episodes of illiquidity in the swap market. To capitalize on the relative advantage of each forecast, we consider a combination of forecasts in a smooth transition framework (as in Chan and Tong (1986) and Teräsvirta (1994)). This smooth transition model allows for time-varying weights to depend upon a transition variable that might indicate when one forecast performs substantially better than another.

Let π_{t+1}^{SU} represent the one-year inflation expectations from the Blue Chip Survey and π_{t+1}^{SW} represent the corresponding one-year expectations from inflation swaps. The smooth transition model assumes two regimes of weighting the forecasts:

$$\pi_{t+1}^{SM} = G(z_t)[\beta_1\pi_{t+1}^{SU} + (1 - \beta_1)\pi_{t+1}^{SW}] + [1 - G(z_t)][\beta_2\pi_{t+1}^{SU} + (1 - \beta_2)\pi_{t+1}^{SW}] + \varepsilon_t,$$

where we assume a logistic transition function

$$G(z_t) = \frac{1}{1 + \exp[-\gamma(z_t - c)]},$$

z_t is the transition variable and $\varepsilon \sim N(0, \sigma^2)$. For regime identification, we restrict $\gamma > 0$ which implies an increase in z_t increases the weight $G(z_t)$ on regime 1. The size of γ determines the speed, or smoothness, of the transition process, with the extremes of $\gamma = \infty$ implying a pure threshold model and $\gamma = 0$ implying no regime-dependence in the model. The threshold parameter c indicates the value at which z_t implies equal weighting across both regimes. We must choose a specific variable z_t that drives the transition function $G(z_t)$. From our previous results, the choice for z_t would optimally capture non-standard periods where liquidity issues drive the inflation swap forecast to send a noisy signal of the market’s true expectation for future inflation, particularly when the swap forecasts substantial deflation. Therefore, our choice for z_t is the lag of the swap forecast, π_t^{SU} . A large decrease in the swap measure could indicate a potential liquidity issue in the inflation swaps market, and relatively more weight should be placed on the survey.¹⁸ In the Appendix, we discuss in detail some alternative transition variables based on the VIX and the previous month’s volatility of changes in inflation compensation.

We estimate the model via the MCMC technique of Gibbs sampling. The parameters are separated in to three blocks: (i) the regime weight coefficients $\beta = [\beta_1, \beta_2]$, (ii) the error variance σ^2 , and (iii) the transition slope γ and threshold c . We construct the full posterior distribution by drawing each parameter block from its conditional posterior distribution given the other blocks. Table 6 lists the prior distributions for the parameters. The draws for the weighting coefficients and variance parameter are standard conditional on the other parameters. We draw $[\gamma, c]$ using the random walk Metropolis-Hastings algorithm as in [Lopes and Salazar \(2006\)](#). We use 10,000 Gibbs iterations after a burn-in period of 20,000 iterations.

Parameter	Prior Distribution	Prior Hyperparameters
β	$N(b_0, B_0)$	$b_0 = [0.5, 0.5]'$, $B_0 = I_2$
σ	$IG(s_0, v_0)$	$s_0 = 0$, $v_0 = 0$
c	$N(c_0, C_0)$	$c_0 = \bar{Z}$, $C_0 = 10$
γ	$\Gamma(g_0, G_0)$	$g_0 = 1$, $G_0 = 0.1$

Table 6: **Prior Distributions for Smooth Transition Model**

Figure 10 shows the weight placed on the swap across time from the smooth transition model.¹⁹ During the early part of the sample, slightly more than two-thirds of the weight is placed on the

¹⁸We considered a number of alternative transition variables, including the gap between the swap and survey forecasts as well as a direct measure of illiquidity in the swap market. In a nontrivial amount of cases, the optimally-weighted forecast when using these transition variables would increase when neither the swap or survey increased in the same period. We found this feature undesirable for a forecast in real-time. Using the level of the swap measure minimized such issues.

¹⁹The weight placed on the swap at any given time period is $G(z_t)(1 - \beta_1) + (1 - G(z_t))(1 - \beta_2)$.

swaps. However, during the episode of illiquidity during the GFC, the weight on swaps drops close to zero. During the subsequent decade-long recovery and expansion, the weight on the swaps fluctuates between half and two-thirds. The weight on swaps temporarily dives during the COVID recession before subsequently rebounding as the economy recovers and the liquidity issues are resolved. During the recent bout of inflation, the weight returns to around 80%, where it began early in the sample. In fact, the swaps receive a majority of the weight in the forecast across the entire sample, with an average weight of about 0.6.

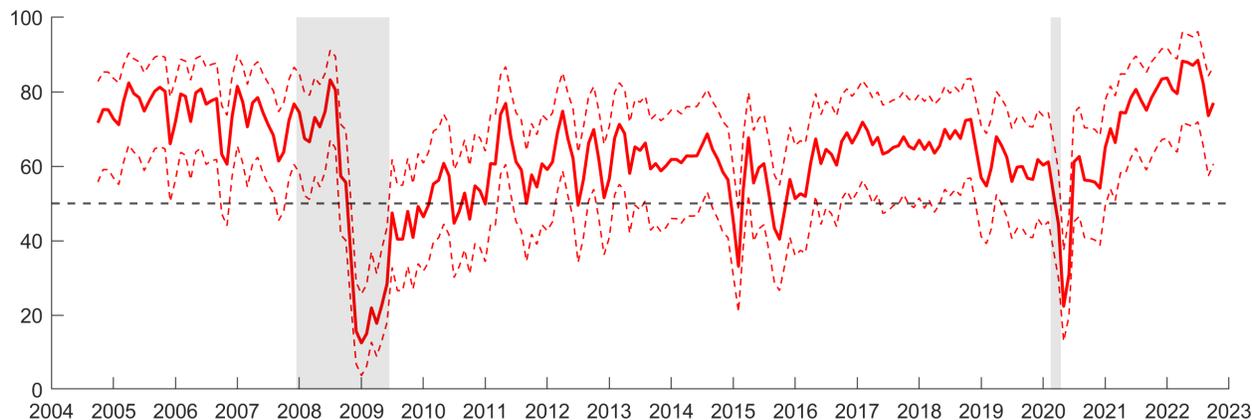


Figure 10: **Smooth Transition Weight on Inflation Swaps**

Note: The solid lines shows the posterior median weight placed on the inflation swap by the smooth transition regime-switching model. The dashed lines show the 68% credible interval. A weight of 100 indicates all weight being optimally placed on inflation swaps. A weight of 0 indicates all weight being optimally placed on surveys. NBER recessions are shown in gray bars.

We present a visual summary of the different possible weights for swaps given their (lagged) level in figure 11. The model gives equal weight to the swaps and survey whenever the swaps are at the estimated threshold $c = 1.00$. When the lagged swap forecasts inflation above this threshold, more weight is given to the swap (and less to the survey). Conversely, the swaps receive less than half the weight as the lagged swaps drop below the threshold. Swaps receive less than 10% of the weight when lagged swaps forecast deflation greater than 1.25%. This result coincides with our previous results that the surveys should receive high weighting when the swaps indicate future deflation (likely because of illiquidity in the inflation swaps market).

We plot the model-based forecast in figure 12, which we call the DCS² measure of inflation expectations.

4.3 Evaluation of Combined Forecasts

In this section, we compare the performance of the forecast from the smooth transition model to the raw inflation swaps and surveys. As in a previous section, we rely upon the CPSA test to diagnose whether there are significant differences in forecasting ability across methods.

Table 7 summarizes the forecast performance for the swaps forecast, the survey forecast, the equally-weighted forecast, and the DCS² forecast from the smooth transition model. The first column shows the mean squared error for each forecast. Despite the high MSE of the swaps, the

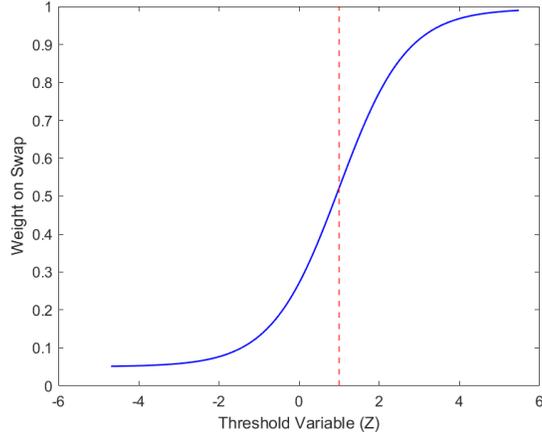


Figure 11: **Conditional Weighting on Inflation Swaps**

Note: The blue line shows the weight placed on the inflation swaps for a given level of the transition variable (the lagged level of the inflation swaps forecast). The red dashed line shows the posterior median for the threshold where the weight is equally distributed across the two regimes.

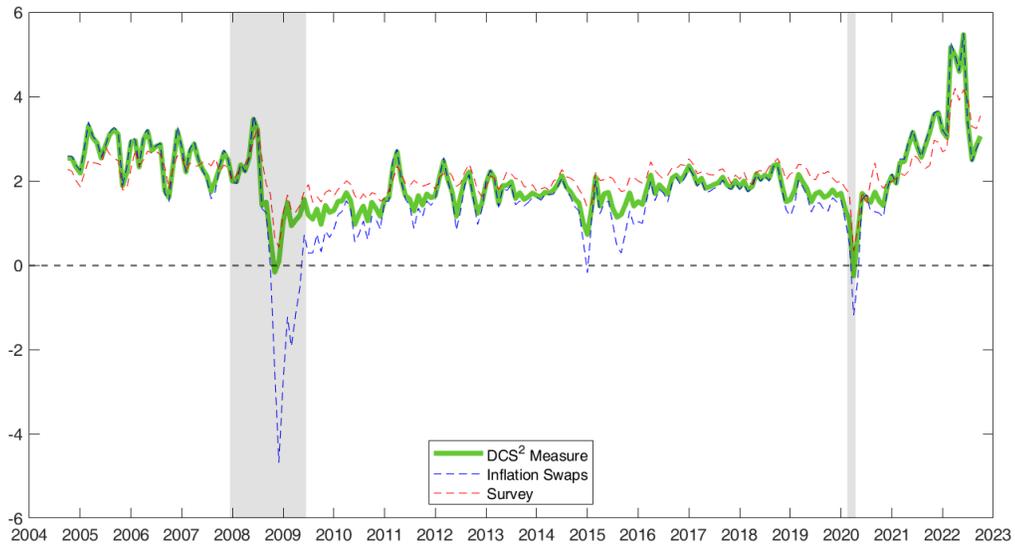


Figure 12: **DCS² Measure of Inflation Expectations**

Note: This figure plots the forecasts from the inflation swaps (dashed blue line), the survey (dashed red line), and the DCS² measure (solid green line). The DCS² measure is the forecast implied by the smooth transition regime-switching model. NBER recessions are shown in gray bars.

DCS² measure achieves the best performance by placing a disproportionate weight on the swaps compared to the surveys, with an average weight of 0.60 on swaps across the entire sample. The second column shows the p-value from a USPA test for each forecast compared to the rest. Recall that the null of the USPA test is that a given benchmark forecast weakly dominates the other three alternative forecasts. We reject the null for both the surveys and the equal-weight forecasts, implying that the best performing alternative forecast dominates these two. Conversely, we fail to reject the null for the inflation swaps and DCS² measure suggesting no alternative forecast performs significantly better than these two approaches across the entire sample on average.

The remaining six columns of table 7 show the p-values for tests of CSPA. We use the same six conditioning variables as we did when comparing the CSPA for the raw swaps and survey measures alone. The survey, swaps, and equal-weight forecasts significantly under-perform compared to the best alternative forecast. However, we cannot reject weak dominance of the DCS² measure for *all* of the conditioning variables at 95% significance. In other words, the DCS² measure is not outperformed by the best alternative forecasting method for any given state of the conditioning variables considered here.

1Y Exp.	MSE	USPA	CSPA						
			INFL	UR	Δ IP	Δ C	Δ RPI	VIX	FFR
Infl. Swaps	4.03	0.105	0.001***	0.001***	0.000***	0.000***	0.001***	0.000***	0.012**
BC Survey	3.93	0.093*	0.000***	0.001***	0.001***	0.000***	0.000***	0.000***	0.003***
Eq. Weights	3.79	0.051*	0.001***	0.006***	0.001***	0.000***	0.000***	0.000***	0.006***
DCS ² Meas.	3.52	1.000	0.998	0.714	0.997	0.964	0.999	0.999	0.999

Table 7: **Significance Tests for Forecasting Ability of All Methods**

Note: The first column shows the mean squared error for the inflation swaps and survey forecasts. The second column shows the p-value for the test for unconditional superior predictive ability (USPA), while the remaining columns show the p-values when testing for conditional superior predictive ability (CSPA). The stars, *, **, ***, denote 90%, 95%, and 99% significance and indicate rejection of the null that a given forecast weakly dominates the alternative forecasts (and, in the case of CSPA, across all states for a given conditioning variable). INFL refers to annual CPI inflation, UR is the unemployment rate, Δ IP is annual growth in industrial production, Δ C is annual growth in personal consumption expenditures, Δ RPI is annual growth in real personal income, VIX is a market-based measure of uncertainty, and FFR is the federal funds rate.

4.4 Out-of-sample evaluation of forecasts

While the inflation swaps, surveys, and equal weighted measure of expectations are model-free and thus not subject to in-sample versus out-of-sample concerns, the results for the smooth transition regime switching model are based on an in-sample analysis. In this section, we conduct the same forecast comparison but only use data up through time t for the smooth regime switching model

to ensure a proper out-of-sample treatment. The training sample for the model is 24 months and the out-of-sample results start in Sept. 2006.

1Y Exp.	MSE	USPA	CSPA						
			INFL	UR	Δ IP	Δ C	Δ RPI	VIX	FFR
Swaps	4.42	0.111	0.001***	0.001***	0.000***	0.001***	0.001***	0.000***	0.007***
Survey	4.24	0.113	0.000***	0.001***	0.006***	0.000***	0.008***	0.001***	0.005***
Eq. Wts.	4.12	0.079*	0.001***	0.009***	0.002***	0.000***	0.000***	0.001***	0.000***
DCS ²	3.87	1.000	0.798	0.764	0.948	0.152	0.999	0.999	0.998

Table 8: **Significance Tests for Out-of-sample Forecasting Ability of All Methods**

Note: The first column shows the mean squared error for the inflation swaps and survey forecasts. The second column shows the p-value for the test for unconditional superior predictive ability (USPA), while the remaining columns show the p-values when testing for conditional superior predictive ability (CSPA). The stars, *, **, ***, denote 90%, 95%, and 99% significance and indicate rejection of the null that a given forecast weakly dominates the alternative forecasts (and, in the case of CSPA, across all states for a given conditioning variable). INFL refers to annual CPI inflation, UR is the unemployment rate, Δ IP is annual growth in industrial production, Δ C is annual growth in personal consumption expenditures, Δ RPI is annual growth in real personal income, VIX is a market-based measure of uncertainty, and FFR is the federal funds rate.

Table 8 shows that the results are little changed when moving to an out-of-sample analysis. The DCS² Measure based on the smooth regime switching model continues to fail to reject the null hypothesis that its forecasts weakly dominate the alternatives. In contrast, we reject the null hypothesis for the rest of the forecasts indicating that there are conditioning states where they are statistically significantly dominated. Overall, these results suggest the DCS² Measure improves over the existing measures when it comes to out-of-sample forecasts and may provide a cleaner signal on one-year inflation expectations.

5 Conclusion

Having a sense of how different measures of expectations of inflation perform in terms of forecasting can be useful for interpreting their signals. Inflation swaps and survey-based measures such as those from Blue Chip are imperfect proxies for the true expectations of financial market participants. While inflation swaps may be subject to risk premiums, surveys suffer from their own set of rigidities and inefficiencies.

On that note, we compare the one-year forecasting efficacy of the two sets of expectations and find that inflation swaps come closer to the realized inflation rate more often but their squared errors are larger on average. We find this apparent disconnect is driven by a few months during the Global Financial Crisis and the early months of the pandemic. When these few periods are excluded, the inflation swaps outperform the surveys across many of the criteria we consider. Moreover, the

inflation swaps outperform the surveys when realized inflation is above its historical median dating back to 2004 and when the gap is more positive than -30 basis points, as has been the case more recently.

We investigate whether combining the swaps and survey measures improves forecasting ability. When using the full sample, we find that a roughly equally-weighted combination is optimal and can outperform the raw measures. Even more weight is allocated to the swaps when optimized over a ten-year rolling window.

Given the relative performance of the swaps and survey across different periods, we utilize a smooth transition regime-switching model to estimate the optimal forecast. The relative weights in the smooth transition model are determined endogenously based on movements in the level of the inflation swaps forecast. The model places most of the weight on the swaps when the inflation swap forecast is above 1.0 percent. However, the model moves weight from the swaps toward the survey when the inflation swap forecasts relatively low inflation or deflation. We name the resulting forecast from this model the DCS² measure of inflation expectations, and show that this measure significantly outperforms the raw swaps and survey measures as well as the equally-weighted forecast. By optimally balancing the signals coming from both sets of forecasts, this approach seems to provide the cleanest measure of one-year inflation expectations. And with these expectations placing relatively more weight on the swaps, which have been lower recently, it suggests that the Federal Reserve is expected to be closer to its inflation target over the next year than the surveys would suggest.

References

- Abrahams, Michael, Tobias Adrian, Richard K Crump, Emanuel Moench, and Rui Yu (2016) “Decomposing real and nominal yield curves,” *Journal of Monetary Economics*, 84, 182–200.
- Ang, Andrew, Geert Bekaert, and Min Wei (2007) “Do macro variables, asset markets, or surveys forecast inflation better?” *Journal of monetary Economics*, 54 (4), 1163–1212.
- Aruoba, Borağan (2020) “Term structures of inflation expectations and real interest rates,” *Journal of Business & Economic Statistics*, 38 (3), 542–553.
- Atkeson, Andrew and Lee E Ohanian (2001) “Are Phillips curves useful for forecasting inflation?” *Federal Reserve bank of Minneapolis quarterly review*, 25 (1), 2–11.
- Bacchetta, Philippe, Elmar Mertens, and Eric Van Wincoop (2009) “Predictability in financial markets: What do survey expectations tell us?” *Journal of International Money and Finance*, 28 (3), 406–426.
- Bauer, Michael D and Erin McCarthy (2015) “Can we rely on market-based inflation forecasts?” *FRBSF Economic Letter*, 30, 1–5.

- Belgrade, Nabyl and Eric Benhamou (2004) “Impact of seasonality in inflation derivatives pricing,” *CDC Ixis Quantitative Research Working Paper No. QRFI*, 08–04.
- Bennett, Julie and Michael Owyang (2022) “On the Relative Performance of Inflation Forecasts,” *Federal Reserve Bank of St. Louis Review*.
- Berge, Travis J (2018) “Understanding survey-based inflation expectations,” *International Journal of Forecasting*, 34 (4), 788–801.
- Campbell, John Y, Robert J Shiller, and Luis M Viceira (2009) “Understanding inflation-indexed bond markets,” Technical report, National Bureau of Economic Research.
- Chan, K. S. and H. Tong (1986) “ON ESTIMATING THRESHOLDS IN AUTOREGRESSIVE MODELS,” *Journal of Time Series Analysis*, 7 (3), 179–190.
- Chen, Ren-Raw, Bo Liu, and Xiaolin Cheng (2010) “Pricing the term structure of inflation risk premia: Theory and evidence from TIPS,” *Journal of Empirical Finance*, 17 (4), 702–721.
- Chernov, Mikhail and Philippe Mueller (2012) “The term structure of inflation expectations,” *Journal of financial economics*, 106 (2), 367–394.
- Christensen, Jens HE, Jose A Lopez, and Glenn D Rudebusch (2010) “Inflation expectations and risk premiums in an arbitrage-free model of nominal and real bond yields,” *Journal of Money, Credit and Banking*, 42, 143–178.
- Chun, Albert Lee and Olfa Maalaoui Chun (2013) “Adjusting Futures Forecasts of Federal Reserve Policy: Risk-Premia or Expectational Errors?”.
- Cieslak, Anna (2018) “Short-rate expectations and unexpected returns in treasury bonds,” *The Review of Financial Studies*, 31 (9), 3265–3306.
- Coibion, Olivier and Yuriy Gorodnichenko (2015) “Is the Phillips curve alive and well after all? Inflation expectations and the missing disinflation,” *American Economic Journal: Macroeconomics*, 7 (1), 197–232.
- Della Corte, Pasquale, Lucio Sarno, and Daniel L Thornton (2008) “The expectation hypothesis of the term structure of very short-term rates: Statistical tests and economic value,” *Journal of Financial Economics*, 89 (1), 158–174.
- Diebold, Francis X and Robert S Mariano (2002) “Comparing Predictive Accuracy,” *Journal of Business & Economic Statistics*, 20 (1), 134–144.
- Diercks, Anthony M and Uri Carl (2019) “A simple macro-finance measure of risk premia in fed funds futures,” *FEDS Notes. Board of Governors of the Federal Reserve System*.
- Diercks, Anthony M and Haitham Jendoubi (2022) “Expectations of Financial Market Participants,” in *Handbook of expectations*, 12, 2–56: Elsevier.

- Diercks, Anthony M, Hiroatsu Tanaka, and Paul Cordova (2021) “Asymmetric Monetary Policy Expectations,” *Available at SSRN 3930267*.
- Diercks, Anthony and Isfar Munir (2020) “Conflicting Signals: Implications of Divergence in Surveys and Market-Based Measures of Policy Expectations.”
- Downing, Chris and Stephen Oliner (2007) “The term structure of commercial paper rates,” *Journal of Financial Economics*, 83 (1), 59–86.
- Durham, J Benson (2003) “Estimates of the term premium on near-dated federal funds futures contracts.”
- D’Acunto, Francesco, Ulrike Malmendier, and Michael Weber (2022) “What Do the Data Tell Us About Inflation Expectations?” Technical report, National Bureau of Economic Research.
- D’Amico, Stefania, Don H Kim, and Min Wei (2018) “Tips from TIPS: the informational content of Treasury Inflation-Protected Security prices,” *Journal of Financial and Quantitative Analysis*, 53 (1), 395–436.
- Elliott, Graham and Allan Timmermann (2013) *Handbook of economic forecasting*: Elsevier.
- Faust, Jon and Jonathan H Wright (2013) “Forecasting inflation,” in *Handbook of economic forecasting*, 2, 2–56: Elsevier.
- Ferrero, Giuseppe and Andrea Nobili (2009) “Futures Contract Rates as Monetary Policy Forecasts,” *International Journal of Central Banking*.
- Fisher, Jonas DM, Chin T Liu, and Ruilin Zhou (2002) “When can we forecast inflation?” *Economic Perspectives-Federal Reserve Bank of Chicago*, 26 (1), 32–44.
- Fleckenstein, Matthias, Francis A Longstaff, and Hanno Lustig (2014) “The TIPS-treasury bond puzzle,” *the Journal of Finance*, 69 (5), 2151–2197.
- Fleming, Michael J and John Sporn (2013a) “How Liquid Is the Inflation Swap Market?” Technical report, Federal Reserve Bank of New York.
- (2013b) “Trading activity and price transparency in the inflation swap market,” *Economic Policy Review*, 19 (1).
- Froot, Kenneth A (1989) “New hope for the expectations hypothesis of the term structure of interest rates,” *The Journal of Finance*, 44 (2), 283–305.
- Fulton, Chad and Kirstin Hubrich (2021) “Forecasting US inflation in real time,” *Econometrics*, 9 (4), 36.
- Giacomini, Raffaella and Halbert White (2006) “Tests of Conditional Predictive Ability,” *Econometrica*, 74 (6), 1545–1578.

- Gil-Alana, Luis, Antonio Moreno, and Fernando Pérez de Gracia (2012) “Exploring survey-based inflation forecasts,” *Journal of Forecasting*, 31 (6), 524–539.
- Gospodinov, Nikolay and Bin Wei (2016) “Forecasts of inflation and interest rates in no-arbitrage affine models,” *Available at SSRN 2730478*.
- Grishchenko, Olesya V and Jing-Zhi Huang (2013) “The inflation risk premium: Evidence from the TIPS market,” *The Journal of Fixed Income*, 22 (4), 5–30.
- Grothe, Magdalena and Aidan Meyler (2018) “Inflation Forecasts: Are Market-Based and Survey-Based Measures Informative?” *International Journal of Financial Research*, 9 (1).
- Hamilton, James D (2009) “Daily changes in fed funds futures prices,” *Journal of Money, credit and Banking*, 41 (4), 567–582.
- Hansen, Peter Reinhard (2005) “A Test for Superior Predictive Ability,” *Journal of Business & Economic Statistics*, 23 (4), 365–380.
- Haubrich, Joseph, George Pennacchi, and Peter Ritchken (2012) “Inflation expectations, real rates, and risk premia: Evidence from inflation swaps,” *The Review of Financial Studies*, 25 (5), 1588–1629.
- Hu, Gang and Mihir Worah (2009) “Why TIPS Real Yields moved significantly higher after the Lehman Bankruptcy,” *PIMCO, Newport Beach, CA*.
- Ichiue, Hibiki and Tomonori Yuyama (2009) “Using survey data to correct the bias in policy expectations extracted from fed funds futures,” *Journal of Money, Credit and Banking*, 41 (8), 1631–1647.
- Kliem, Martin and Alexander Meyer-Gohde (2022) “(Un) expected monetary policy shocks and term premia,” *Journal of Applied Econometrics*, 37 (3), 477–499.
- Kliesen, Kevin L (2015) “How accurate are measures of long-term inflation expectations?” *Economic Synopses* (9).
- Li, Jia, Zhipeng Liao, and Rogier Quaadvlieg (2021) “Conditional Superior Predictive Ability,” *The Review of Economic Studies*, 89 (2), 843–875.
- Lloyd, Simon P (2020) “Overnight Indexed Swap-Implied Interest Rate Expectations,” *Finance Research Letters*, 101430.
- Longstaff, Francis A (2000) “The term structure of very short-term rates: New evidence for the expectations hypothesis,” *Journal of Financial Economics*, 58 (3), 397–415.
- Lopes, Hedibert F. and Esther Salazar (2006) “Bayesian Model Uncertainty In Smooth Transition Autoregressions,” *Journal of Time Series Analysis*, 27 (1), 99–117.

- Martinez, Andrew B (2020) “Extracting Information from Different Expectations,” Technical report.
- Pflueger, Carolin E and Luis M Viceira (2011a) *An empirical decomposition of risk and liquidity in nominal and inflation-indexed government bonds*: National Bureau of Economic Research.
- (2011b) “Inflation-indexed bonds and the expectations hypothesis,” *Annu. Rev. Financ. Econ.*, 3 (1), 139–158.
- Sack, Brian (2004) “Extracting the expected path of monetary policy from futures rates,” *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 24 (8), 733–754.
- Schmeling, Maik, Andreas Schrimpf, and Sigurd Steffensen (2020) “Monetary policy expectation errors,” *Available at SSRN 3553496*.
- Stock, James H and Mark W Watson (2007) “Why has US inflation become harder to forecast?” *Journal of Money, Credit and banking*, 39, 3–33.
- (2010) “Modeling inflation after the crisis,” Technical report, National Bureau of Economic Research.
- Teräsvirta, Timo (1994) “Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models,” *Journal of the American Statistical Association*, 89 (425), 208–218, [10.1080/01621459.1994.10476462](https://doi.org/10.1080/01621459.1994.10476462).
- Thomas, Lloyd B (1999) “Survey measures of expected US inflation,” *Journal of Economic perspectives*, 13 (4), 125–144.
- To, Thuy Duong and Ngoc-Khanh Tran (2019) “Cheap TIPS or expensive inflation swaps? Mispricing in real asset markets,” *Mispricing in Real Asset Markets (January 28, 2019)*.
- Trehan, Bharat (2015) “Survey measures of expected inflation and the inflation process,” *Journal of Money, Credit and Banking*, 47 (1), 207–222.
- Verbrugge, Randal J and Saeed Zaman (2021) “Whose Inflation Expectations Best Predict Inflation?” *Economic Commentary* (2021-19).
- Wang, Zhenyu and Wei Yang (2018) “OIS Risk Premiums.”
- White, Halbert (2000) “A Reality Check for Data Snooping,” *Econometrica*, 68 (5), 1097–1126.

Appendix

A Comparison to the Random Walk

To be clear, the focus of this study is on forecasts that inherently reflect inflation expectations. There are numerous specifications and approaches that can likely provide superior forecasts, but in our view, these specifications have little to say about the true expectations because they do not inherently reflect them. For instance, a superior forecast may increase over a given time period, but if the surveys and inflation swap rates move in the opposite direction, it becomes difficult to suggest the superior forecast says anything about the true expectations.

Nonetheless, we can compare the expectations in this study to a common forecast such as the random walk. The random walk inflation rate was shown by [Atkeson and Ohanian \(2001\)](#) to be very difficult to beat many years ago and consisted of using the previous observed inflation rate as the current forecast. [Table A.1](#) shows the results when we include the random walk. Across all of our metrics, we see that the random walk has the highest errors. This suggests that while it may have been difficult to beat prior to 2000, the most recent sample suggests it is not as good of a performer. There are numerous other approaches we could compare in terms of forecast performance, but we view this as outside the scope of this study.

		No Deflation
One-Year Expectation	Full Sample	No Deflation
Panel A: Mean Squared Error		
Inflation Swaps	4.03	3.68
Blue Chip Survey	3.93	3.99
Random Walk	6.18	5.60

Table A.1: **Comparison with Random Walk Inflation Expectation**

Note: Full sample is from September 2004 to October 2022 (one-year realized inflation did not exist beyond this date at the time of writing). The No Deflation sample omits the 12 months out of 206 in which the inflation swaps projected deflation.

B Inflation Swaps vs TIPS Breakevens

In this section, we provide a direct comparison between the inflation swaps and the TIPS breakevens. [Figure B.1](#) shows the monthly difference between the inflation swaps and TIPS breakevens (positive values indicating the swaps are larger).²⁰ One can see that the largest differences occur during

²⁰For this exercise, we use a seasonally-adjusted one-year inflation expectation coming from TIPS. See [Belgrade and Benhamou \(2004\)](#) for discussion of seasonality in pricing of inflation securities.

episodes of low liquidity around 2008 and beginning of 2020. This potentially provides some additional evidence that the inflation swaps are relatively less subject to liquidity issues. Consistent with the tables above which show that the inflation swaps outperform the TIPS breakevens, when we put the two measures against each other in the optimal weighting exercise based on a rolling 10-year window, we see the optimal weight is close to 100% on the inflation swaps for most of the sample in Figure B.2. This result suggests that the inflation swaps may have provided a cleaner signal about future inflation compared to TIPS based measures.

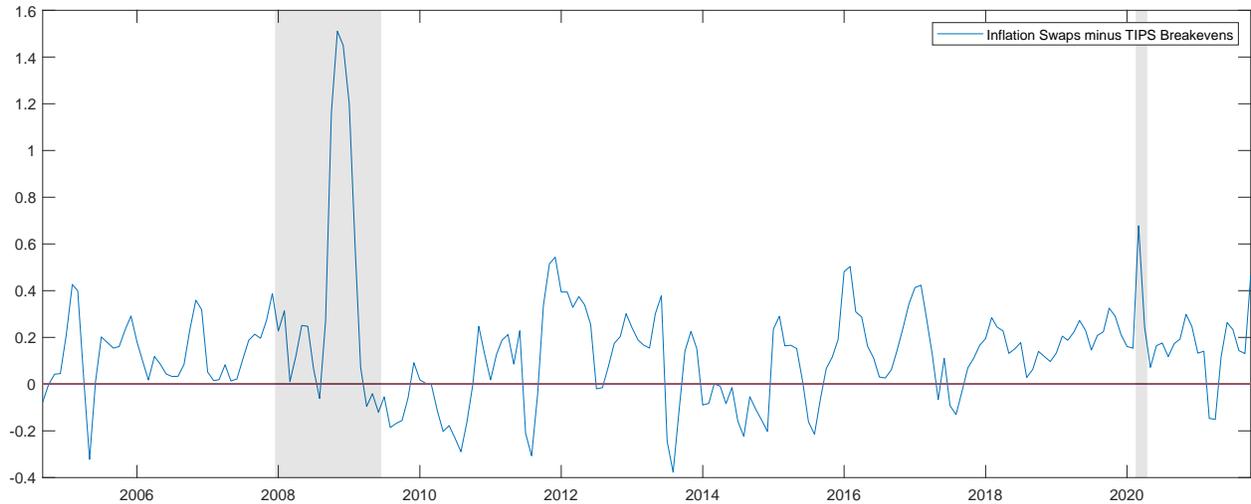


Figure B.1: Difference between Inflation Swaps and TIPS Breakevens

Note: The difference represents inflation swaps minus TIPS breakevens. Gray bars denote NBER recessions. Source: Blue Chip; Barclays; Bloomberg.

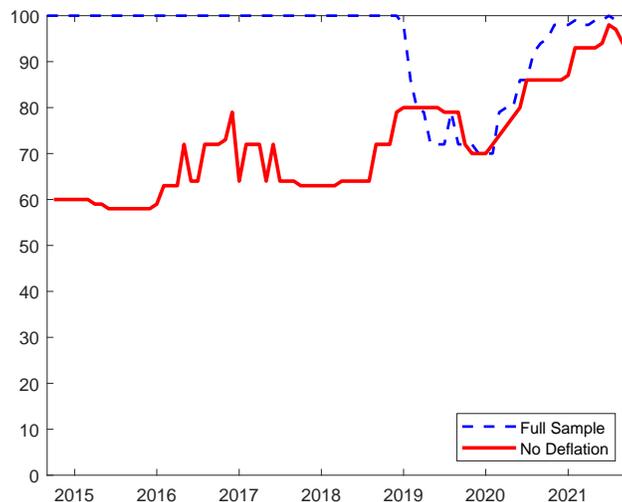
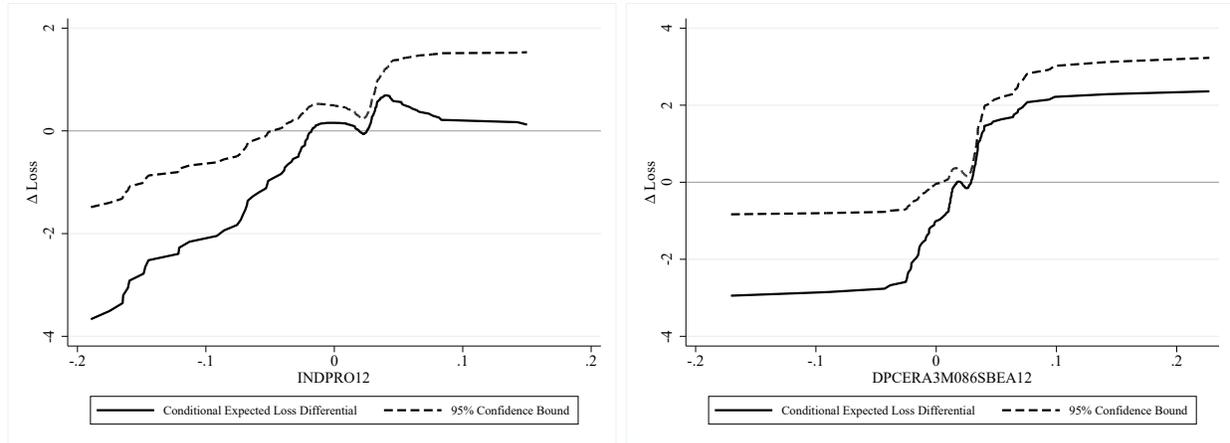


Figure B.2: Optimal Inflation Swap vs TIPS Breakevens Weights: Minimum Absolute Error, Rolling 10-Year Window

Note: A weight of 100 implies all weight being optimally placed on the inflation swaps. Source: Blue Chip; Barclays; Bloomberg.

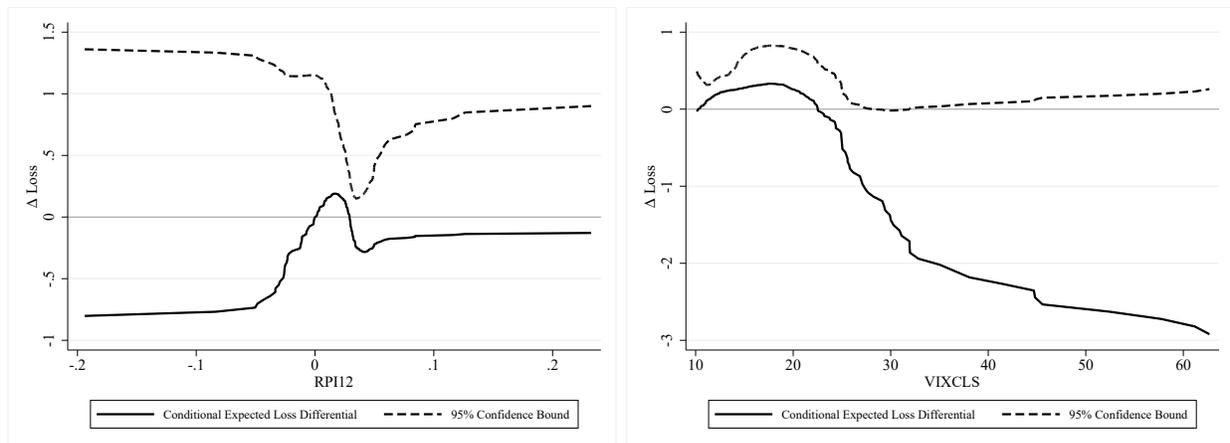
C CSPA Conditional Expected Loss Differentials

In this section, we plot the expected conditional loss differentials for each of the conditioning variables that we test for in the main text.



**Figure C.1: CSPA Loss Differentials for Survey minus Swaps:
Industrial Production Growth (left), PCE Consumption Growth (right)**

Note: This figure shows the expected loss differential (survey squared error minus swap squared error) across the conditioning variables of Industrial Production Growth and the PCE Consumption Growth .



**Figure C.2: CSPA Loss Differentials for Survey minus Swaps:
Real Personal Income Growth (left), Vix (right)**

Note: This figure shows the expected loss differential (survey squared error minus swap squared error) across the conditioning variables of Real Personal Income Growth and the Vix.

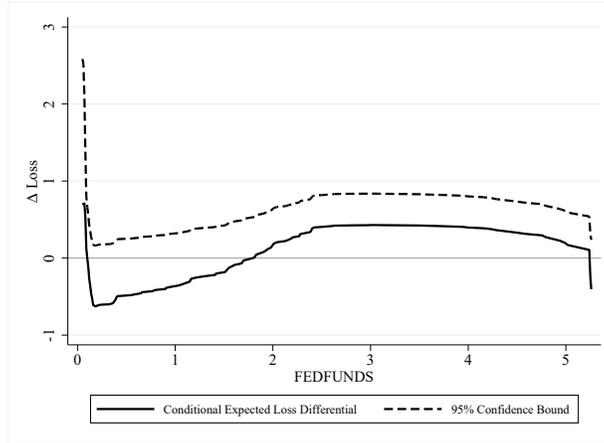


Figure C.3: CSPA Loss Differentials for Survey minus Swaps:
Federal Funds Rate

Note: This figure shows the expected loss differential (survey squared error minus swap squared error) across the conditioning variable of the federal funds rate.

C.1 No Deflation Sample

In this section we plot the same figures but for the No Deflation sample.

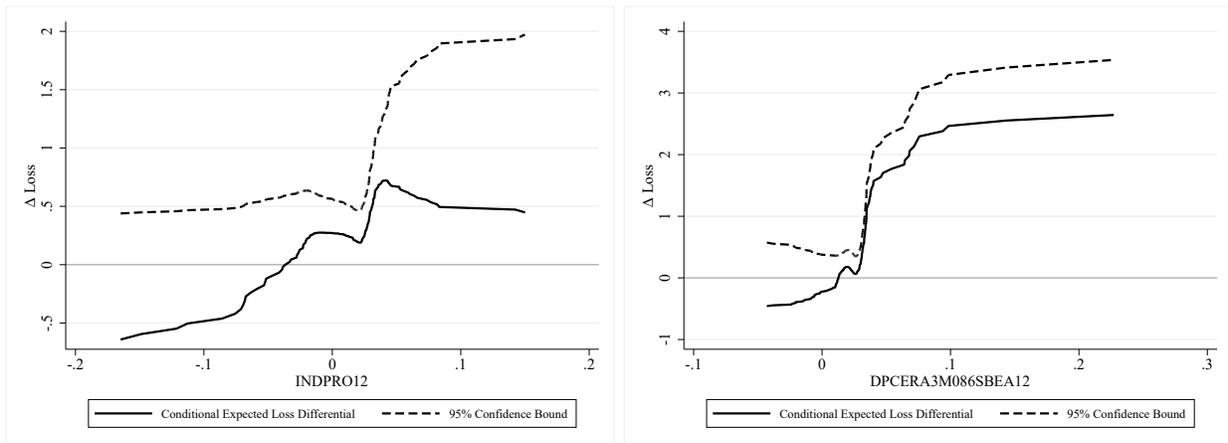


Figure C.4: CSPA Loss Differentials for Survey minus Swaps:
Industrial Production Growth (left), PCE Consumption Growth (right)

Note: This figure shows the expected loss differential (survey squared error minus swap squared error) across the conditioning variables of Industrial Production Growth and the PCE Consumption Growth .

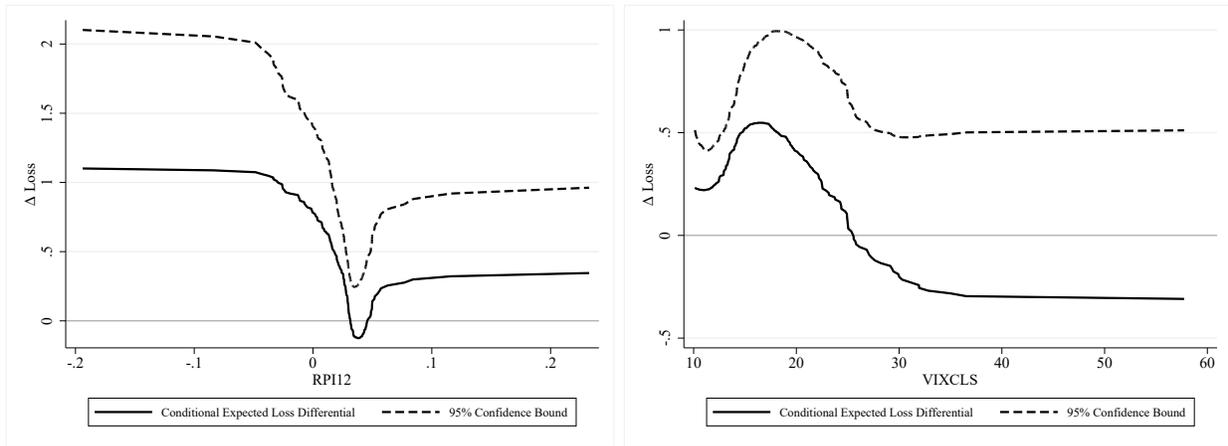


Figure C.5: CSPA Loss Differentials for Survey minus Swaps:
Real Personal Income Growth (left), Vix (right)

Note: This figure shows the expected loss differential (survey squared error minus swap squared error) across the conditioning variables of Real Personal Income Growth and the Vix.

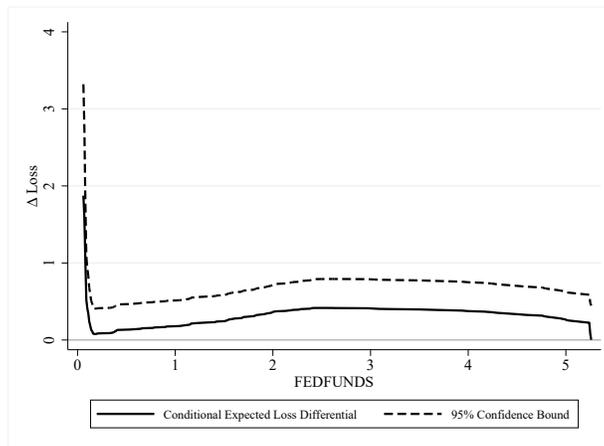


Figure C.6: CSPA Loss Differentials for Survey minus Swaps:
Federal Funds Rate

Note: This figure shows the expected loss differential (survey squared error minus swap squared error) across the conditioning variable of the federal funds rate.

C.2 Out-of-Sample Comparison with Smooth Regime Switching Model

In this section we plot the same figures but use the smooth regime switching model as the benchmark to compare to the rest of the models. Again, if the expected conditional loss is significantly below zero, it indicates another forecast's squared error loss is lower than the smooth regime switching model's for that state of the world.

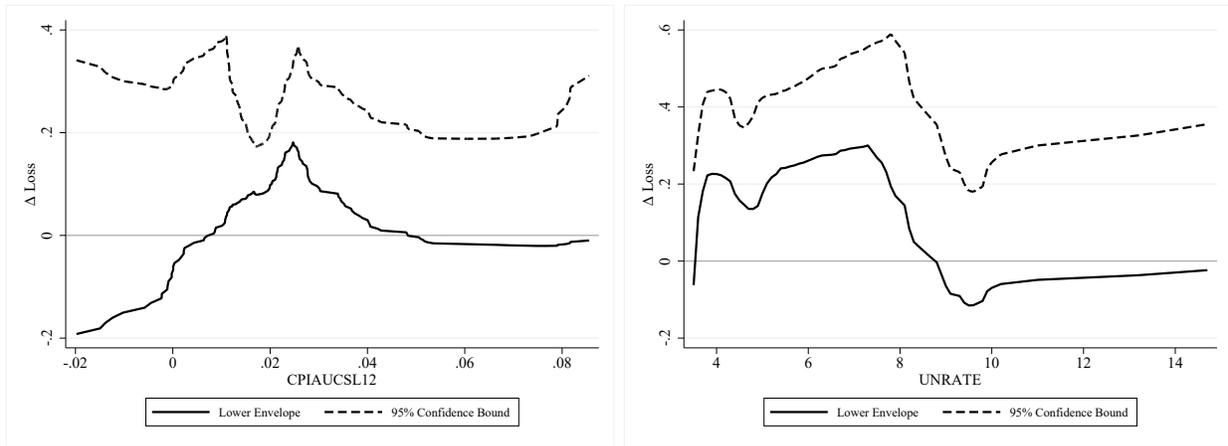


Figure C.7: CSPA Loss Differentials for Smooth Regime Switching vs All: Inflation (left), Unemployment (right)

Note: This figure shows the expected loss differential (alternative forecast minus smooth regime switching) across the conditioning variables of inflation and the unemployment rate.

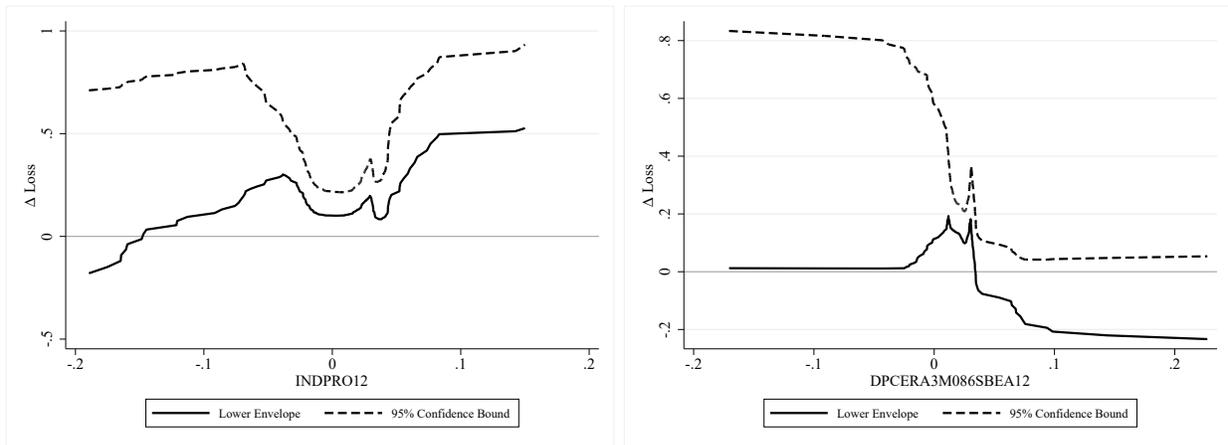


Figure C.8: CSPA Loss Differentials for Smooth Regime Switching vs All: Industrial Production Growth (left), PCE Consumption Growth (right)

Note: This figure shows the expected loss differential (alternative forecast minus smooth regime switching) across the conditioning variables of Industrial Production Growth and the PCE Consumption Growth .

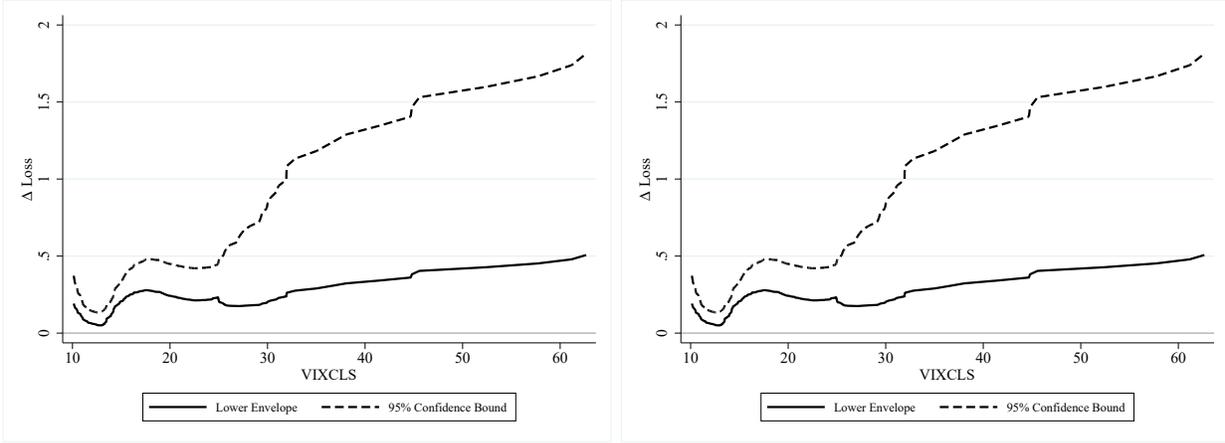


Figure C.9: **CSPA Loss Differentials for Smooth Regime Switching vs All: Real Personal Income Growth (left), Vix (right)**

Note: This figure shows the expected loss differential (alternative forecast minus smooth regime switching) across the conditioning variables of Real Personal Income Growth and the Vix.

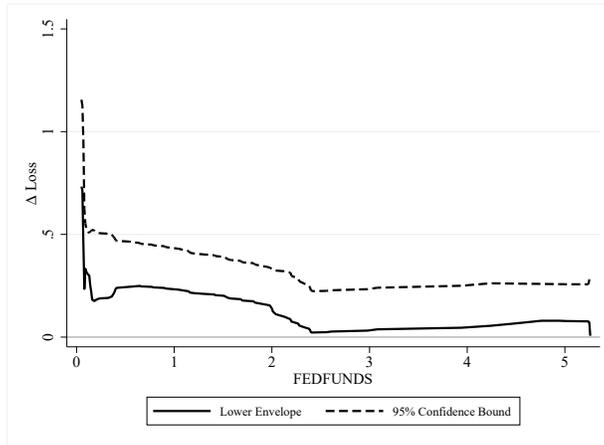


Figure C.10: **CSPA Loss Differentials for Smooth Regime Switching vs All: Federal Funds Rate**

Note: This figure shows the expected loss differential (alternative forecast minus smooth regime switching) across the conditioning variable of the federal funds rate.

D Alternative Transition Variables

In the main text, the benchmark smooth regime switching model’s transition variable is the level of the inflation swap. In this section, we explore how our results change if we use alternative transition variables based on the VIX and the monthly volatility of the daily changes in inflation compensation (specifically it’s Z-score).²¹

²¹For the inflation compensation, we use the 5-year TIPS-based inflation compensation as this measure has relatively better liquidity than shorter alternative horizons and thus could provide a cleaner proxy.

Both the VIX and the monthly volatility of the daily changes in inflation compensation seem like good candidates as they largely comove quite closely with the level of the inflation swap. For instance, during both the global financial crisis and the pandemic, the VIX and the monthly vol. of inflation compensation both jumped considerably, while the level of the inflation swap dramatically declined. This can be seen in the panels below in Figure D.1.

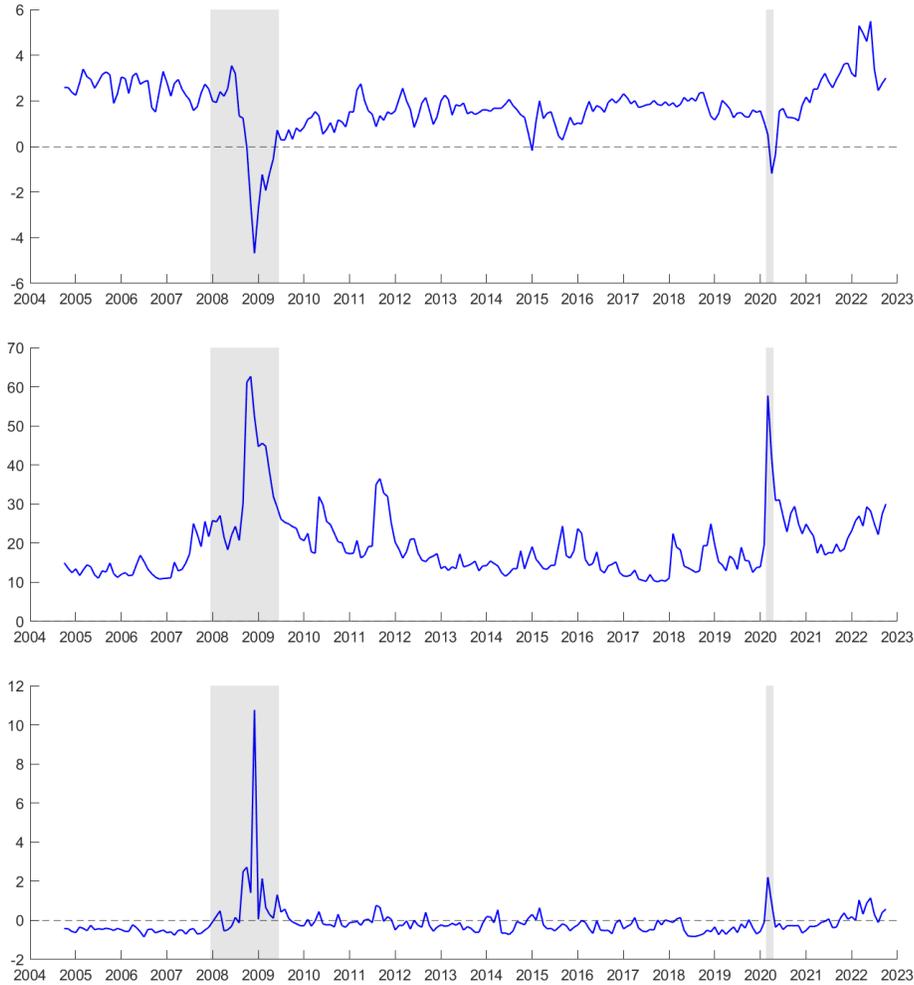


Figure D.1: Swap level (top), VIX (middle) Vol. of changes in infl. comp. (bottom)

Note: This figure shows a time series of the three different transition variables used in our analysis. The top panel shows the inflation swap level, the middle panel shows the VIX, and the bottom panel shows the volatility of daily changes in inflation compensation.

In terms of the weights on the inflation swap, Figure D.2 shows that for each case, the relative weights on the swap decline, or put differently, the weight on the surveys rise during the two previous recessions. A key difference though is that for the post-pandemic period, the swap level transition variable puts higher weight on the swaps while the VIX and monthly vol.-based transition variables put relatively less weight on the swaps. In addition, the VIX transition variable seems to be putting considerable weight on the swaps during a substantial portion of the sample. This

is in contrast to the vol. and level of swap transition variables, which more frequently put their weight on the swaps around 60-70%. The reason for this difference could be due to the lack of meaningful variability in the vol. of changes in infl. comp as can be seen from roughly 2010 to 2020. In contrast, the VIX seems to have more variation over that time period that also correlates well with the outperformance of the swaps and thus, the VIX version tends to place higher weight on the swaps.

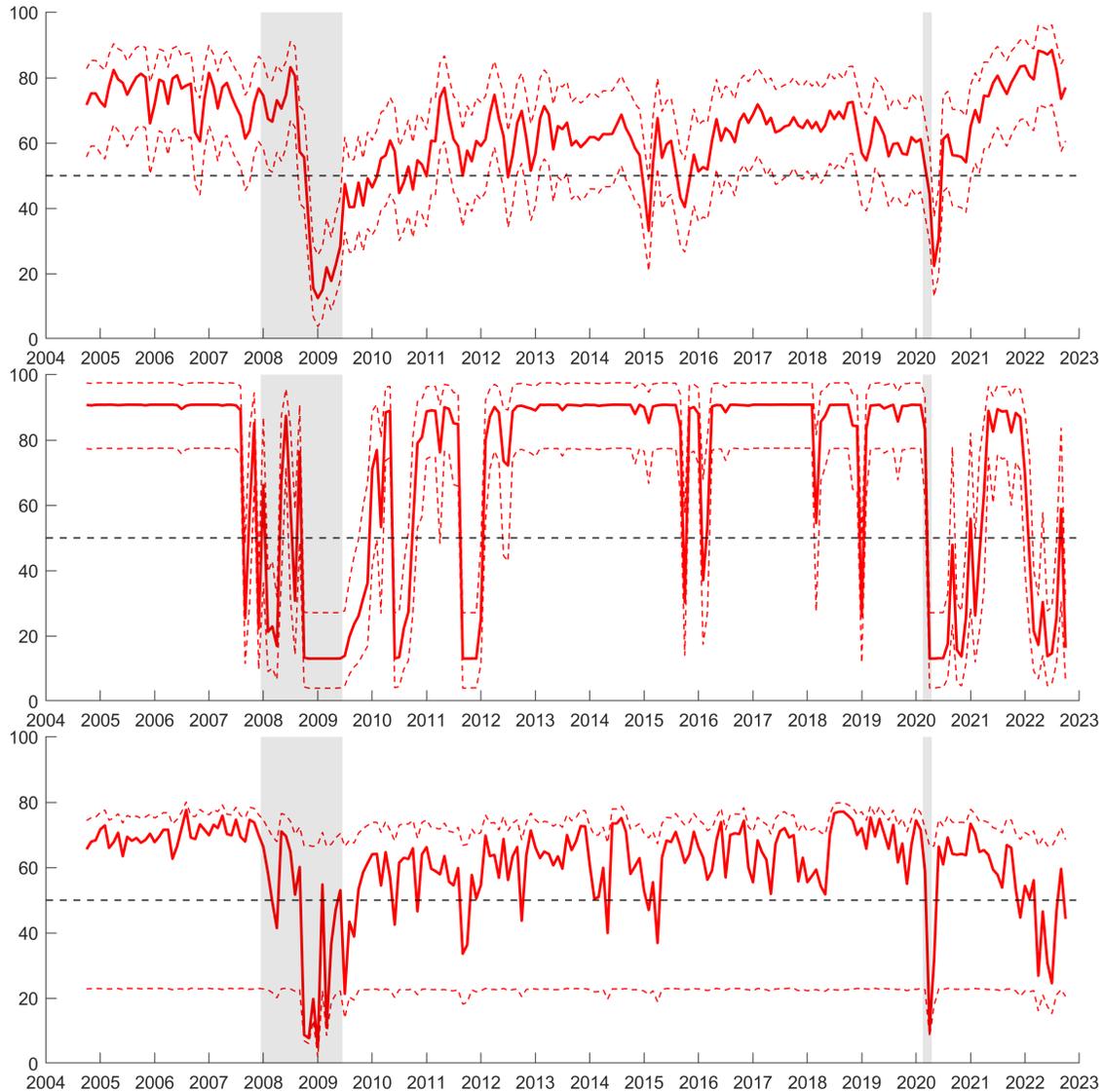


Figure D.2: **Smooth Transition Weights on Swap based on Swap level (top), VIX (middle) Vol. of changes in infl. comp. (bottom)**

Note: The solid lines shows the posterior median weight placed on the inflation swap by the smooth transition regime-switching model. The dashed lines show the 68% credible interval. A weight of 100 indicates all weight being optimally placed on inflation swaps. A weight of 0 indicates all weight being optimally placed on surveys. NBER recessions are shown in gray bars.

For completeness, we also plot the logistic functions along with the transition variables as follows in Figure D.3.

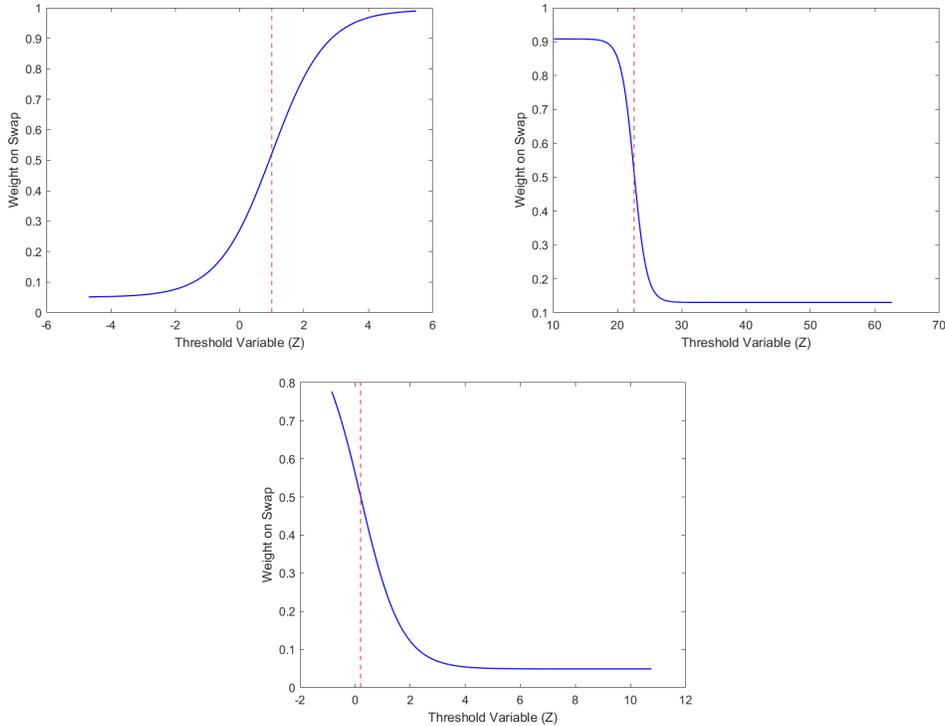


Figure D.3: **Conditional Weighting on Inflation Swaps based on Swap level (top left), VIX (top right) Vol. of changes in infl. comp. (bottom)**

Note: The blue line shows the weight placed on the inflation swaps for a given level of the transition variable. The red dashed line shows the posterior median for the threshold where the weight is equally distributed across the two regimes.

Lastly, we explore the out-of-sample forecast performance of the three measures in Table D.1. We can see that the DCS² Measure typically has the highest p-value indicating the highest failure to reject the null that it is weakly superior to the other approaches. The VIX-based approach has fairly similar performance as the benchmark but does have two conditioning variables where the CSPA test rejects the null of weak superiority. In contrast, the vol-based measure seems to do the worst among the three measures.

To better understand why the vol-based measure is not doing as well, we can plot the conditional loss function against the conditioning states as we did in the main text. Recall that the conditional expected loss function is the competing measures minus the squared error of the vol-based measure. If the competing measures have lower squared errors for a given state, the conditional expected loss will be negative at those states. Figure D.4 shows that the vol-based measure doesn't do as well when inflation was high and the unemployment rate was low post-pandemic. That's because with the vol-based measure being slightly elevated, the model was putting less weight on the swap, which led to worse performance than the benchmark's level based measure, which increased the weight on the swap during this time period.

1Y Exp.	MSE	USPA	CSPA						
			INFL	UR	ΔIP	ΔC	ΔRPI	VIX	FFR
Vol-based	4.17	0.084*	0.001***	0.002***	0.004***	0.000***	0.005***	0.002***	0.001***
VIX-based	4.00	0.277	0.306	0.286	0.026**	0.448	0.039**	0.445	0.154
DCS ²	3.87	1.000	0.975	0.978	0.950	0.995	0.158	0.961	0.904

Table D.1: **Significance Tests for Out-of-sample Forecasting Ability of All Methods**

Note: The first column shows the mean squared error for the three different transition variables based on the volatility of the previous month's changes in inflation compensation, the approach based on the VIX, and the benchmark approach using the level of the swap. The second column shows the p-value for the test for unconditional superior predictive ability (USPA), while the remaining columns show the p-values when testing for conditional superior predictive ability (CSPA). The stars, *, **, ***, denote 90%, 95%, and 99% significance and indicate rejection of the null that a given forecast weakly dominates the alternative forecasts (and, in the case of CSPA, across all states for a given conditioning variable). INFL refers to annual CPI inflation, UR is the unemployment rate, ΔIP is annual growth in industrial production, ΔC is annual growth in personal consumption expenditures, ΔRPI is annual growth in real personal income, VIX is a market-based measure of uncertainty, and FFR is the federal funds rate.

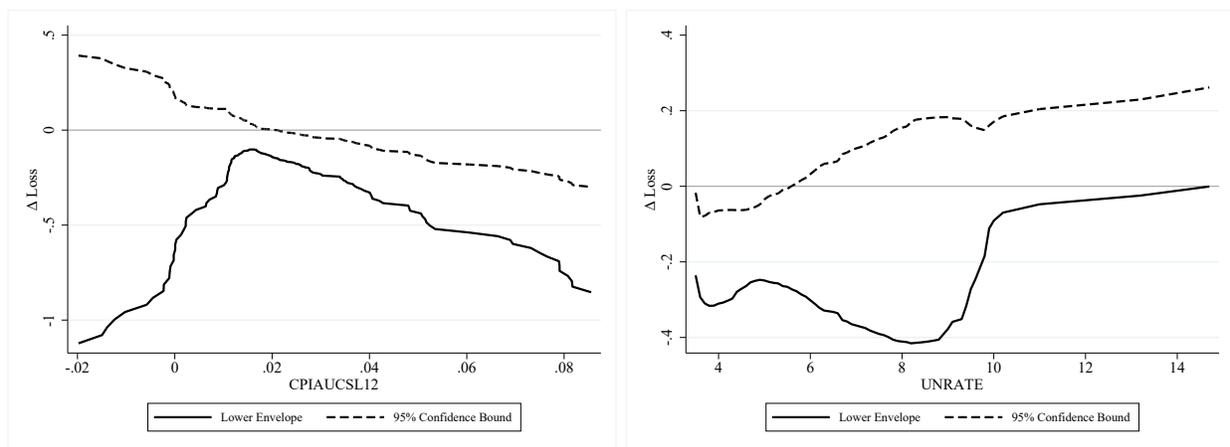


Figure D.4: **CSPA Loss Differentials for Vol-based Smooth Reg. Switch. Model vs All: Realized Inflation (left) and Unemployment Rate (right)**

Note: This figure shows the expected loss differential (alternative forecast minus vol-based smooth regime switching model) across the conditioning variables of realized inflation and the unemployment rate.